

مطبوعات
مكتبة الملك فهد الوطنية
السلسلة الثانية
(٧٨)

نظم استرجاع قواعد بيانات الويب غير المرئية

دراسة تحليلية لوضع مواصفات محركات البحث

د. سيد ربيع سيد إبراهيم

الرياض
١٤٣١هـ / ٢٠١٠م

نظم استرجاع قواعد بيانات الويب غير المرئية

دراسة تحليلية لوضع مواصفات محركات البحث

Invisible Web Databases Retrieval Systems

An analytical study to put a suggested search engines characteristics

د. سيد ربيع سيد إبراهيم

مدرس بقسم المكتبات والوثائق

كلية الآداب – جامعة بني سويف

مكتبة الملك فهد الوطنية

الرياض : ١٤٣١هـ/ ٢٠١٠م

ح مكتبة الملك فهد الوطنية ، ١٤٣١هـ

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

إبراهيم، سيد ربيع سيد

نظم استرجاع قواعد بيانات الويب غير المرئية : دراسة تحليلية لوضع مواصفات محركات البحث/ سيد ربيع سيد إبراهيم .- الرياض، ١٤٣١هـ

٤١٥ص؛ ٢٤سم .- (السلسلة الثانية؛ ٧٨)

ردمك : ٣-٣٧٣-٠٠-٩٩٦٠-٩٧٨

١- تخزين واسترجاع المعلومات ٢- الإنترنت أ. العنوان ب. السلسلة

ديوي : ٠٠٤/٦٧ ١٤٣١/٩٠٧٩

رقم الإيداع : ١٤٣١/٩٠٧٩

ردمك : ٣-٣٧٣-٠٠-٩٩٦٠-٩٧٨

مكتبة الملك فهد الوطنية

ص.ب : ٧٥٧٢

الرياض : ١١٤٧٢

هـاتف : ٤٦٢٤٨٨٨

فاكس : ٤٦٤٥٣٤١

البريد الإلكتروني : info@kfnl.gov.sa

الإهداء

إلى هؤلاء الناس...

أم . . عليَّ بكل جوارحها أعطت وحنّت،

والد . . يداه لي بالجود مُدّت،

إخوة . . قلوبهم عليَّ بالحب حنّت،

زوجة . . في كامل حياتها رقّت ما تخلّت،

ابنة . . لإكمال سعادتي هلّت

أساتذة . . ألبابهم وقلوبهم لي ربّت . .

إليهم جميعًا أهدي عملي

المحتويات

المقـــدمــــــة

أولاً: تمهيد :

باتت الإنترنت تشكل المستودع الرئيسي للمعرفة البشرية المسجلة، التي اعتمدت في اتصالها وتواصلها واختزان معلوماتها وموادها على الشكل الرقمي. وكما اختلفت مصادر المعلومات في الشكل الرقمي بين النصية والمصورة والصوتية، فقد اختلفت أيضًا في حجم وطريقة الوصول إليها ؛ حيث أفرزت الإنترنت صفحات الويب ذات الروابط الفائقة html وملفات الصور والصوت إلى جانب ظهور عدد لا نهائي من قواعد البيانات العلمية والمتخصصة، وتحمل هذه القواعد معلومات ذات قيمة تعلو غيرها من صفحات الويب لإخضاعها لتحكيم وإشراف الهيئات العلمية المختلفة.

ويمكن القول إنه كما يزداد فضاء الكون اتساعًا بالأجرام السماوية، فقد نمت أعداد قواعد البيانات على الويب إلى الدرجة التي أصبح حجم هذه القواعد يفوق حجم صفحات وملفات الويب الأخرى بما يقدر بخمسمائة ٥٠٠ مرة. إلى جانب اختلاف هذه القواعد في طبيعة بنيتها وتنظيمها وطرق الوصول إلى معلوماتها عن طبيعة تنظيم وبحث صفحات الويب الأخرى؛ ولهذا التضخم في الحجم والاختلاف في الطبيعة فقد شكلت هذه القواعد جزءاً منفصلاً عن الويب الحالية أطلق عليه عدة مصطلحات هي (الويب الخفية والويب غير المرئية والويب العميقة) وهي على التوالي Hidden web, Invisible web, Deep web. ويوما بعد يوم شكلت الويب غير المرئية وليدا جديدا للويب يحتاج إلى رعاية أكثر حرصًا ودأبًا على تحقيق تنظيمه وتسهيل الوصول إلى ما يحويه من معلومات محجوبة عن مستخدمي الويب ومحركات البحث الآنية.

وقد شكلت الويب غير المرئية (قواعد البيانات والمواقع التفاعلية) حاجزًا منيعًا أمام محركات بحث الويب الحالية في التعامل مع ملفات تلك القواعد، لأن هذه القواعد تعمل على إتاحة مصادرها من خلال واجهات البحث search interfaces الخاصة بها، إلى جانب اختلاف احتواء هذه القواعد على الكثير من أنواع

الملفات الرقمية مثل pdf, doc, ppt. فضلاً عن خصوصية الملكية الفكرية التي تتميز بها قواعد البيانات عن غيرها من مصادر المعلومات على الويب، كما أن قواعد بيانات الويب غير المرئية تحتاج إلى مستويات أكثر تعمقًا في التحليل من برامج الزاحف التي قلما تتعدى المستوى الثاني أو الثالث في تحليل روابط صفحات الويب URLs. وعلى ذلك فإننا بصدد إما إهمال هذا الكم من الملفات الرقمية عالية القيمة لباحثي الويب أو التعامل مع هذه الظاهرة بتطوير أدوات العمل والتحليل لمصادر المعلومات الرقمية على الويب، بما يمكن معه مسايرة هذا الجزء المتطور في الويب. ولقد ذكرت إحدى الدراسات إحصاءات حول الحجم المعلوماتي المتضخم لقواعد بيانات الويب غير المرئية كما يلي(1):

- تحوي الويب غير المرئية حجمًا هائلاً من المعلومات الرقمية يصل إلى ٧٥٠٠ تيرا بايت، مقارنة بالويب التقليدية التي تحوي فقط حجم ١٩ تيرا بايت من المعلومات الرقمية(❖).
- تحوي الويب العميقة ما يقرب من ٥٥٠ بليون وثيقة إلكترونية مفردة مقارنة بالويب السطحية التي تحوي فقط واحد بليون وثيقة، وذلك فقط في عام ٢٠٠١م.
- يوجد الآن أكثر من ٢٠٠,٠٠٠ موقع داخل الويب غير المرئية.
- أن الويب غير المرئية تأخذ في النمو السريع قياسًا بما يحدث في بيئة الويب المرئية؛ حيث أوضحت إحدى الإحصائيات أن حجم كشاف محرك البحث Google قد قفز في مايو ٢٠٠٤ من ٣,٣ مليار صفحة إلى ٤,٢ مليار صفحة ويب.
- يحوي ٦٠ موقعًا فقط من مواقع الويب غير المرئية على ٧٥٠ تيرا بايت من المعلومات الرقمية.

---

(١) Deep Web Search Tools. Educational Web Portal at BHSU , 2004cited cited at 12/2/2006. http://www.bhsu.edu/education/edfaculty/lturner/Deep%20Web%20Search%20Tools1.htm.

(❖) التيرابايت : احدى وحدات قياس السعة التخزينية في الحاسب الآلي، وتساوي ١٠٢٤ GB جيجا بايت.

- تحتوي الويب غير المرئية على معلومات تصل نسبتها إلى ١٠٠٠ – ٢٠٠٠ بالمقارنة مع قيمة معلومات الويب التقليدية.

تشكل قواعد بيانات الويب غير المرئية كمًا هائلاً من مصادر المعومات التي لا يمكن لمؤسسات المكتبات أو مراكز المعلومات العمل بدونها؛ حيث يعني كشف وإتاحة هذه القواعد على الويب مصدرًا جديدًا يضاف إلى مقتنيات المكتبات من المصادر الإلكترونية على الخط المباشر. وثمة جانب آخر لا يمكن إغفاله وهو المكتبات الرقمية التي تضع قواعد بياناتها على الويب، ويتزايد عدد هذه القواعد بشكل سريع وفي وقت قصير، وكونها تبقى معزولة عن برامج الزاحف فهي بعيدة عن تكشيف محركات البحث وبعيدة بدورها عن مستخدمي الويب، هذا في الوقت الذي تسعى المكتبات فيه إلى الوصول إلى المستفيدين في أماكنهم. فضلاً عن قواعد البيانات التي تحمل مصادر معلومات متباينة الأشكال مثل قواعد بيانات المصادر الرقمية المصورة الثابتة والمتحركة العامة أو العلمية، قواعد بيانات مصادر المعلومات الصوتية ، قواعد الهيئات المتحفية والطبية والعلمية ... وغيرها.

ثانيًا: مشكلة الدراسة:

أدى التقدم المستمر للحاسبات الآلية في جانبي الرقمنة Digitization والتصغير Minimization على حد سواء إلى تحول الكم الأعظم من صفحات الويب إلى اتخاذ شكل قواعد البيانات ، وذلك لكي تتمكن أكثر من التعامل مع مستخدمي الويب إلى جانب اختزان كميات هائلة من البيانات داخل قواعد بياناتها. على الجانب الآخر فإن قواعد البيانات العلمية التي تحوي النصوص الكاملة أو الملفات الرقمية في أشكال متباينة مثل الأكروبات والنصوص pdf, doc. Xle, ppt... وغيرها. وهذه القواعد تشكل بذاتها مصادر المعلومات ذات القيمة الأعلى في فضاء الويب خلاف صفحات الويب الأخرى.

وتكمن مشكلة الدراسة في أن الأعداد المليونية من قواعد البيانات على الويب لا يستطيع الباحثون الوصول إليها عن طريق أدوات البحث العادية باستخدام محركات أو أدلة البحث ؛ وذلك يرجع إلى عجز برامج الزحف داخل محركات البحث عن التعامل مع قواعد بيانات الويب لتعدد أنواع الملفات داخلها أو عجز هذه البرامج عن تحقيق مستوى التعمق اللازم لبحث تلك القواعد أو قيود وحقوق الملكية التي تضعها قواعد البيانات أمام الزاحف قبل الوصول إلى مصادرها. فضلاً عن اختلاف طبيعة البناء والتنظيم والبحث من قاعدة بيانات إلى أخرى مما يتطلب تدخل العنصر البشري في إضافة وبحث الويب غير المرئية. وقد دفعت هذه الحيثيات إلى صناعة ظاهرة توقف عندها الباحث في الكيفية اللازمة للتعامل مع بحث وتنظيم مصادر قواعد بيانات الويب غير المرئية، إلى جانب فتح أبواب تلك القواعد أمام مستخدمي الويب من خلال نوافذ محركات بحث الويب. وما معايير التعامل التي يجب إصباغ برامج الزحف بها لتمكينها من الوصول إلى قلب تلك القواعد من مصادر المعلومات؟ وما التغييرات اللازمة لمحركات بحث الويب لتمكينها من رؤية الجانب الخفي من الويب؟ وعلى ذلك فقد اعتمدت هذه الدراسة في صياغة المشكلة البحثية على الفروض التالية:

- لا تستطيع محركات بحث الويب بنمطها الحالي التعامل مع قواعد بيانات الويب ذات الأشكال المختلفة من مصادر المعلومات.
- إن اختلاف طبيعة بناء قواعد البيانات (الويب غير المرئية) عن صفحات الويب html (الويب التقليدية) يحتاج بدوره إلى جيل متطور من محركات البحث بمعايير أكثر تطورًا.

ثالثًا: أسئلة الدراسة:

تتلخص مشكلة الدراسة في مجموعة الأسئلة التالية :

١- ما الملامح الخاصة للويب غير المرئية، التي تشكل من خلالها بيئة جديدة خلاف بيئة الويب المرئية؟

٢- ما الملامح الكمية والكيفية التي تُختص بها قواعد البيانات غير المرئية على الويب؟

٣- ما أساليب التنظيم والمعالجة الفنية لمحتوى قواعد بيانات الويب غير المرئية؟

٤- ما متطلبات الجيل القادم من محركات البحث القادر على تنظيم وبحث محتوى الويب غير المرئية؟

٥- ما الأساليب الأكثر ملاءمة للتعامل مع محتوى قواعد البيانات مباشرة من جانب المستفيدين، بما يحقق الوصول إلى معلومات عالية القيمة العلمية أمام المستفيدين؟

٦- ما ملامح العلاقة والتعامل بين أدوات بحث الويب وملفات قواعد البيانات؟

٧- ما مواصفات ومعايير عمل برامج الزاحف والمكشف داخل محركات بحث الويب غير المرئية، التي تستطيع من خلالها الوصول إلى محتوى قواعد بيانات الويب غير المرئية وإخضاعه للبحث من جانب المستفيدين؟

رابعًا: تعريفات إجرائية:

- **الويب:**

تعد الويب أحد مكونات شبكة الإنترنت ، وهي مجموعة الصفحات أو المواقع التي تتميز باستخدام النص الفائق، والوسائط المتعددة داخل صفحاتها ذات بنية النصوص المعيارية الفائقة.

- **الويب المرئية:**

هي مجموعة صفحات أو مواقع الويب ذات القابلية للتنظيم والبحث والاستدعاء بواسطة محركات البحث، حيث لا تعتمد هذه الصفحات على قواعد بيانات لتخزين ملفاتها أو معلوماتها.

- **الويب غير المرئية:**

هي مجموعة الصفحات أو مواقع الويب التي تعتمد على قواعد بيانات مختلفة الحجم ، ولذلك لا تستطيع محركات البحث التعامل معها؛ لعدم قدرة برنامج الزاحف على التعامل مع آليات التنظيم والبحث في قواعد بياناتها.

- **قواعد البيانات:**

هي ملف أو اكثر من ملفات البيانات التي يتم تنظيمها في شكل تسجيلات يمكن استخدامها والرجوع إليها إذا دعت الحاجة الموضوعية. وتعتمد قاعدة بيانات الويب على العنصر البشري في بناء وتسيير ملفات قاعدة البيانات. وبطبيعتها تعتمد قواعد البيانات على آليات تختلف عن تلك المستخدمة داخل محركات بحث الويب.

وتصنف قواعد البيانات بحسب طبيعة ما تحتوي عليه من بيانات ؛ فتوجد قواعد البيانات الببليوجرافية كقواعد الفهرسة المقروءة آليًا MARC مارك ، وقواعد بيانات الكشافات مثل MEDLINE في الطب، وقواعد بيانات المستخلصات مثل LISA في المكتبات والمعلومات. إلى جانب قواعد البيانات غير الببليوجرافية مثل Library literature . وتتبنى هذه الدراسة قواعد البيانات غير الببليوجرافية التي تحتوي على ملفات النصوص الكاملة.

- **محركات البحث:**

هي برامج آلية لتنظيم واسترجاع صفحات أو مواقع الويب، معتمدة على برامج فرعية مثل الزاحف والمكشف لأداء مهام محددة. وتؤدي محركات البحث عملها بشكل تام مع مواقع الويب غير أنها لا تستطيع أداء مهامها بشكل دقيق داخل قواعد البيانات التي تحكمها آليات عمل مختلفة.

- **أدلة البحث الموضوعية:**

هي برامج لتنظيم وبحث صفحات أو مواقع الويب معتمدة على استخدام العنصر البشري في تكشيف وتنظيم هذه المواقع؛ ولاستخدامها العنصر البشري

فإن أدلة البحث تتعامل مع قواعد البيانات بشكل اكثر كفاءة من تعامل محركات البحث مع قواعد البيانات. مثل دليل YAHOO!.

خامسًا : أهداف الدراسة :

تحاول هذه الدراسة تحقيق مجموعة من الأهداف المتعلقة بتنظيم ومعالجة قواعد بيانات الويب مثل:

١- التعرف إلى ملامح بيئة جديدة من الويب داخل الإنترنت تندرج تحت اسم "الويب غير المرئية"

٢- إظهار الملامح الكمية والإحصائية والكيفية التي تتواجد بها قواعد البيانات غير المرئية على الويب.

٣- فهم الطبيعة الخاصة لقواعد بيانات الويب واختلافها عن غيرها من مصادر الويب الأخرى.

٤- دراسة أساليب التنظيم والمعالجة لملفات المعلومات المختلفة داخل قواعد بيانات الويب.

٥- التعرف إلى متطلبات الجيل القادم من محركات البحث القادرة على تنظيم وبحث مختلف أشكال ملفات المعلومات داخل بيئة الويب.

٦- طرح أساليب متطورة للتعامل مع قواعد البيانات مباشرة من جانب المستفيدين، مما يحقق الوصول إلى معلومات عالية القيمة بالنسبة المستفيدين.

٧- التعرف إلى ملامح العلاقة والتعامل بين أدوات بحث الويب وملفات قواعد البيانات.

٨- وضع مواصفات وملامح جديدة لمحركات بحث الويب ، تستطيع من خلالها إخضاع محتوى قواعد البيانات إلى التنظيم والبحث من جانب مستخدمي الويب.

سادسًا : أهمية الدراسة :

تأتي أهمية هذه الدراسة من العوامل التالية :

١- ينشغل مجتمع المعلومات في هذه الآونة بفك رموز التعامل مع ملفات الويب والتي تدور في ببيئة html ؛ وذلك لكونها سهلة الوصول صعبة المعالجة. غير أن هذه الصعوبة ستزداد أكثر في حالة ملفات الويب غير المرئية وهي ملفات صعبة الوصول صعبة المعالجة.

٢- إن تجاهل التعامل مع هذا الكم من المعلومات يشكل في حد ذاته ضياع قدر كبير من المعلومات يصل إلى ٥٥٠ ضعف الكم المتداول من المعلومات الرقمية بين محركات وأدلة بحث الويب.

٣- تحتوي قواعد بيانات الويب غير المرئية على أكثر ملفات المعلومات قيمة التي تتبع في معظمها هيئات علمية تفرض المنهجية والضبط والتحكيم على معلوماتها البحثية.

٤- تناول الإنتاج الفكري الأجنبي ظاهرة الويب غير المرئية بالوصف فقط دون التطرق الملائم لهذه الظاهرة بالدراسة والتحليل. أما العالم العربي فلم يرد به ما يذكر عن هذه الظاهرة. وذلك وفقًا لمحاولات بحث على مواقع الأطروحات العلمية مثل Dissertation Abstracts International.

٥- إن الإحصاءات سالفة الذكر عن حجم ومواقع الويب غير المرئية تجعل من الضروري التنبه إلى وجود بيئة جديدة تمامًا تشكل عناصرها قواعد البيانات التي تحتاج إلى نظرة أكثر تطورًا في التعامل مع الملفات الإلكترونية تنظيمًا وبحثًا.

٦- ما زالت أدوات بحث الويب تعمل جاهدة على التطوير والتحديث لحل مشكلات التنظيم والاسترجاع لملفات الويب المرئية html ، مما يعكس

درجة التأخر في الانتباه إلى صياغة ملامح أكثر تطورًا للتعامل مع بيئة قواعد بيانات الويب غير المرئية.

٧- القصور المرتبط باستخدام البرامج الآلية داخل محركات البحث مثل الزاحف والمكشف؛ حيث لا يمتلك برنامج الزاحف إلى الآن القدرة على التعامل مع محتويات قواعد البيانات، مما يعني قصور محركات البحث في التعامل مع حجم هائل ضائع من معلومات الويب.

سابعًا: حدود الدراسة:

تسعى الدراسة إلى تحقيق أهدافها من خلال الحدود التالية:

- **الحدود الموضوعية:**

تركز الدراسة على موضوع المعالجة والتنظيم والاسترجاع لملفات قواعد بيانات الويب غير المرئية من خلال محركات بحث قواعد بيانات الويب العميقة. إلى جانب دراسة مواصفات أدوات البحث الأكثر ملاءمة للتعامل مع بيئة قواعد البيانات غير المرئية.

- **الحدود اللغوية:**

تتناول الدراسة قواعد البيانات غير المرئية ومحركات البحث التي تحمل ملفات باللغة الإنجليزية.

- **الحدود المكانية:**

تتناول الدراسة المعالجة والتنظيم لقواعد بيانات الويب غير المرئية الموجودة على مواقع شبكة الإنترنت.

- **الحدود النوعية:**

تتناول الدراسة محركات بحث قواعد البيانات غير المرئية التي تحمل أنواع الملفات النصية والوسائط المتعددة، مثل ملفات الصوت والفيديو وملفات النصوص.

- **الحدود الزمنية:**

عمل الباحث على قراءة ملامح قواعد بيانات ومحركات بحث الويب غير المرئية حتى الآن، بالإضافة إلى إجراء الدراسة التطبيقية في الفترة بين ٢٠٠٨/١٠/٢٠ و ٢٠٠٨/١٢/١ مع الاهتمام بتحديث نتائجها حتى ٢٠٠٩/٣/١م.

ثامنًا : منهجية الدراسة :

اعتمدت الدراسة على استخدام منهجي البحث الوصفي التحليلي والتجريبي، وذلك كما يلي :

أولاً: المنهج الوصفي التحليلي ؛ وذلك لرصد ومتابعة الآليات التي تعمل بها قواعد البيانات في معالجة ملفاتها الرقمية، وأيضًا أساليب عمل محركات بحث قواعد البيانات في التعامل مع تلك الملفات الرقمية وذلك من خلال رصد الإنتاج الفكري حول هذه الظاهرة ومن ثم وصفها وتفسيرها، حتى يتم الخروج باستنتاجات ذات دلالة موضوعية لمشكلة البحث.

وقد اعتمدت الدراسة في **جانبها النظري** على استعراض أدبيات الموضوع للتعرف على جوانبه المختلفة فيما يتعلق بقواعد البيانات وتكوينها وشكل ومعالجة الملفات الرقمية داخلها، إلى جانب ما يتعلق بمحركات بحث قواعد البيانات وخصوصية عمل هذه المحركات.

ثانيًا: المنهج التجريبي؛ وذلك لملاحظة ومشاهدة واقع العمل بين محركات البحث وقواعد البيانات بالاعتماد على عينة قصدية من محركات وقواعد بيانات الويب غير المرئية.

أدوات البحث :

١- أداة الملاحظة والمعايشة مع الإنترنت؛ من خلال مشاهدة ومراقبة أساليب البحث والاسترجاع في محركات بحث قواعد البيانات، إلى جانب اختبار أساليب العمل داخل قواعد البيانات ذاتها.

٢- **أدوات الدراسة التجريبية** : وقد انقسمت إلى قسمين هما :

القسم الأول : التجهيزات المادية : حيث اعتمد الباحث في الجزء التطبيقي من الدراسة على مجموعة ثابتة من التجهيزات التي استخدمت بشكل ثابت أثناء التعامل مع عينة الدراسة من قواعد بيانات ومحركات بحث الويب غير المرئية، وهي على النحو التالي :

١- جهاز حاسب آلي IBM ThinkPad بالإمكانات التالية:

- معالج Intel 1.2 GHz.
- ذاكرة رئيسة 256 MB of RAM.
- لوحة أساسية FIC.
- كارت FAX Modem.
- شاشة "14 LG.

٢- خط هاتف بخاصية Dial up بسرعة 56 Kb.

القسم الثاني : مصطلحات اختبار الدراسة التطبيقية :

ركزت الدراسة اهتمامها على آليات وملامح الأداء في كل من قواعد بيانات الويب، ومحركات بحث الويب غير المرئية، دون التركيز على طبيعة الموضوعات أو المجالات موضع البحث. مما أدى إلى اختيار مصطلحات بحث عامة غير محددة بموضوع أو سمات خاصة، وذلك عند اختبار آليات التنظيم والمعالجة داخل محركات بحث الويب غير المرئية. أما في حالة الدراسة التطبيقية لقواعد البيانات فقد عمل الباحث على محاولة اختيار مصطلحات البحث بما يتوافق مع التخصص الموضوعي أو النوعي لقواعد البيانات. وذلك على النحو التالي :

الجدول رقم (٠-١) مصطلحات اختبار الدراسة التطبيقية

عنصر الاختبار	مصطلحات الاختبار
محركات بحث الويب غير المرئية	information retrieval systems, information retrieval, earth
قواعد بيانات الموسيقى والأفلام	Music, Amr diab
قواعد بيانات السيرة	Opama, Zeweel
قواعد بيانات الطب	Blood flow , Heart
قواعد بيانات علم الحيوان	Lion , dogs

عينة الدراسة :

- تعتمد الدراسة على عينة قصدية من قواعد بيانات الويب غير المرئية ومحركات بحث تلك القواعد، واعتمد اختيار العينة على ما جاء من ترشيح لأهم قواعد وأدوات بحث الويب غير المرئية في South Central Regional Library Council بولاية New York لاختيار قواعد بيانات الويب الخفية والمحركات البحث بها(❖). وقد جاء اختيار العينة اعتماداً على الخصائص المتشابهة في بناء قواعد البيانات التي تحمل أشكالاً مختلفة من الملفات على الويب. إلى جانب ذلك فقد اعتمد الباحث على اختيار أهم أدوات البحث المرشحة من قبل التقرير السنوي لأبحاث الويب غير المرئية[1] Deep Web Research 2006 (❖). وهي كالتالي:

(❖) http://www.lakenet.org/net_ref/manuals/invisible.html

(١) http://www.llrx.com/features/deepweb2006.htm, cited at 1/3/2006.

(❖) اعتمد الباحث في اختيار العينة وترجى عناصرها في تحليل أهم مواقع الحديث عن الويب غير المرئية وأكثر أدوات البحث ذكراً وأهمية في هذه الواقع. ولمزيد من القراءة يمكن الرجوع إلى المواقع التالية:

- http://www.llrx.com/features/deepweb2006htm
- http://www.closerlooksearch.com
- http://www.weblens.org/invisible.html

أولاً: قواعد بيانات الويب غير المرئية: وتتمثل في نوعين هما :

الجدول (٠- ٢) عينة قواعد بيانات الويب غير المرئية

التخصص	قواعد متخصصة الموضوع	قواعد وسائط متعددة
السير الذاتية	http://www.biography.com/ ***	www.corbis.com
الحيوان	http://animalsearch.net/	www.freefoto.com
التربية	http://www.eric.ed.gov/ ***	http://www.rhapsody.com
الطب	http://medlineplus.gov/	http://www.allmusic.com/
التسويق الإلكتروني	http://www.amazon.com/***	http://movietome.com/
		http://www.world-of-movies.com/

ثانيًا: أدوات بحث الويب غير المرئية: تتمثل هذه الأدوات في نماذج محركات وأدلة البحث وهي كالتالي:

الجدول (٠- ٣) عينة أدوات بحث الويب غير المرئية

أدلة البحث لقواعد بيانات الويب غير المرئية	محركات بحث متخصصة للويب غير المرئية
http://www.completeplanet.com/	http://turbo10.com
www.magportal.com	http://infomine.ucr.edu
www.findarticles.com	http://www.incywincy.com/
http://www.genealogyspot.com/	http://business.exploritnow.com/
http://www.beaucoup.com/	http://www.scienceresearch.com/search/
http://www.academicinfo.net/	http://www.scirus.com/srsapp/

خطوات الدراسة:

انتهجت الدراسة مجموعة من الخطوات نوضحها كما يلي:

أولاً: تعمل الدراسة في بداية العمل على تحديد الأطر النظرية والاستقرائية لنشأة وتطور الويب غير المرئية. من خلال إرساء المفاهيم الأساسية لمكونات وطبيعة الويب غير المرئية.

ثانيًا: تتناول الدراسة الواقع الفعلي لقواعد بيانات الويب غير المرئية؛ وذلك من حيث أنواع المحتويات بها، آليات العمل داخل قواعد البيانات، وآليات التنظيم والمعالجة ثم الأساليب المستخدمة لاستدعاء محتوياتها من الملفات الرقمية مختلفة الأشكال.

ثالثًا: التعرف إلى طبيعة أدوات بحث الويب غير المرئية من محركات وأدلة ؛ وذلك بتناول طبيعة وآليات التعامل داخل كل أداة والتعرف إلى عمليات المعالجة من تكشيف وتصنيف لمحتويات قواعد بيانات الويب غير المرئية. وعلى ذلك يمكن العمل على دراسة ملامح استخدام العنصر البشري وبرامج الزاحف داخل أدوات البحث للتعامل مع محتويات قواعد البيانات.

رابعًا: تتناول الدراسة الواقع التطبيقي لآليات تنظيم وبحث محتويات قواعد بيانات الويب غير المرئية داخل أداتي المحركات وأدلة البحث ، من خلال دراسة وتحليل نتائج التطبيق المعتمد على مجموعة من الخصائص سابقة التحديد.

خامسًا: تنتهي الدراسة عند وضع المواصفات المقترحة لما يمكن أن تكون عليه أدوات معالجة وبحث قواعد بيانات الويب غير المرئية على الإنترنت.

تاسعًا: الدراسات السابقة:

استعرض الباحث أدبيات الموضوع ووجد أن الإنتاج الفكري العربي لم يتناول قواعد بيانات الويب غير المرئية، سواء كان ذلك في الأطروحات الأكاديمية أو الدراسات البحثية. غير أن الإنتاج الفكري الأجنبي عمل في العامين الأخيرين على الانتقال من الحديث عن هذا الموضوع إلى مرحلة الكتابات القصيرة إلى الدراسات المعمقة المتمثلة في الأطروحات الأكاديمية. وكان أهم ما أفرزه الإنتاج الفكري الأجنبي في هذا الموضوع عن طريق بحث مستخلصات الرسائل الجامعية International Abstracts Dissertation Digital (UMI)Proquest[1] ، ومع

(١) http://wwwlib.umi.com/dissertations/, cited a 1/3/2006.

استخدام مصطلحات مثل ... Invisible web databases, deep web وأفرز ذلك الدراسة التالية:

Ipeirotis, Panagiotis G. Classifying and searching hidden-web text databases , advisor: Prof. Luis Gravano , Columbia University, 2004 (١).

لقد عملت هذه الدراسة على استقراء محتوى قواعد البيانات النصية فقط، محاولة تحديد التقسيمات الموضوعية المهمة لمحتوى تلك القواعد . كما اعتمدت في الإجراءات على برنامج QProber لبحث واستكشاف محتوى قواعد البيانات النصية. وباستطلاع نتائج الاستفسارات البحثية الموجهة من هذا البرنامج إلى محتوى قاعدة البيانات، يستطيع البرنامج تحديد أهم الكلمات المفتاحية الدالة المستخدمة في محتوى هذه القاعدة. ويتخذ هذا البرنامج أسلوبين لتحديد الكلمات الدالة داخل محتوى قاعدة البيانات: الأسلوب الأول : التعامل إحصائيًا مع مجموعة النتائج المستدعاة لتحديد أكثر الكلمات المتعلقة بقاعدة البيانات. أما الأسلوب الثاني : فهو تحديد بنية النتائج الهرمية للاستدلال على موضوعات محتوى قاعدة البيانات.

غير أن الدراسة الحالية تختلف عن هذه الدراسة في النواحي التالية:

- اقتصرت الدراسة السابقة على التعامل مع المحتوى النصي، من حيث الاستدعاء والتحليل، متجنبة تمامًا التعامل مع قواعد البيانات غير النصية ؛ حيث لا يستطيع البرنامج التعامل مع metadata ميتاداتا تلك الملفات.
- لم تتطرق الدراسة السابقة إلى استخدام أدوات بحث الويب (محركات وأدلة البحث) ؛ حيث اقتصر الحديث عن المتابعة البشرية لمجريات عمل برنامج Qprober داخل قواعد البيانات.

(١) Ipeirotis, Panagiotis G. Classifying and Searching Hidden-Web Text Databases , Advisor: Prof. Luis Gravano , Columbia University, 2004. http://wwwlib.umi.com/dissertations

- لم تهتم الدراسة السابقة بتحديث آليات عمل أدوات بحث الويب لملاءمة التعامل مع مصدر قواعد بيانات الويب كمصدر للمعلومات. وإخضاع محتوى هذا المصدر للبحث والاستدعاء أمام مستخدمي الويب كباقي أشكال مصادر معلومات الويب.

وهذه الجوانب هي ما تسعى الدراسة الحالية لتحقيقه.

عاشراً: قراءة في أدبيات الدراسة:

تمثل قضية الويب غير المرئية وإمكانات تنظيم واسترجاع المحتوى غير المرئي، قضية جديدة على ساحة نظم استرجاع المعلومات الرقمية. ومن ثم فإن اعتماد الدراسة على هذا الكم من المصادر وأدبيات دراسة هذه الظاهرة، قد أوجد عند الباحث شغفاً كبيرًا لدراسة ملامح الإنتاج الفكري لهذه الظاهرة، وما استندت عليه من كتابات في بدايات الحديث عنها وتناولها. وقد اعتمد الباحث في هذه القراءة على استخدام أسلوب تحليل الاستشهادات المتمثلة في مصادر بناء هذه الدراسة، وقد استثنى الباحث من مصادر الدراسة مواقع الويب، ليعتمد فقط على المقالات والدراسات التي تتناول بالشرح والتأصيل ظاهرة الويب غير المرئية. وقد بلغ العدد الإجمالي للمصادر المعتمد عليها ١١٠ مصادر.

وتأتي أهمية هذه القراءة من وجهة نظر الباحث في أنها تمثل جوانب اهتمام أخرى للويب غير المرئية؛ حيث يمكن من خلال قراءة أعداد المصادر المستشهد بها ودلالتها التعرف إلى نتائج مختلفة، ما كانت لتصل إليها هذه الدراسة من خلال نصوص المصادر ذاتها. وقد عمل الباحث في قراءة أدبيات الدراسة على الكشف عن ماهية جوانب محددة في معالجة مشكلة الويب غير المرئية، تكمن هذه الجوانب فيما يلي :

أولاً: التأصيل الدلالي والاصطلاحي للظاهرة.

ثانيًا: اتجاهات حل مشكلة الويب غير المرئية.

ثالثًا: أداة البحث الأقرب للمشكلة.

رابعًا: التركيز على معالجة برامج الزاحف.

خامسًا: التوزيع الزمني لمصادر الدراسة.

سادسًا: التوزيع الموضوعي لمصادر الدراسة.

سابعًا: التوزيع الجغرافي واللغوي لمصادر الدراسة.

أولاً: التأصيل الدلالي للظاهرة (بين الخفية وغير المرئية والعميقة):

إن المفاهيم دائمًا ما ترتبط بالمصطلحات الدالة عليها، وإذا ما كان هناك مفهوم ذا دلالة محددة لا يمكن أن يبقى ويصمد إذا لم يوضع في قالب اصطلاحي مخصص. من ثم فقد حاول الباحث تتبع المصطلح أو المصطلحات التي تدل على جوهر ظاهرة الويب غير المرئية، وقد لاحظ الباحث أن هناك مصطلحات عدة تستخدم كمترادفات للتعبير عن طبيعة ظاهرة الدراسة، هذه المصطلحات هي الويب غير المرئية، والويب العميقة، والويب الخفية، والويب غير الشفافة، والويب الخاصة. غير أن طبيعة استخدام هذه المصطلحات قد كشف أن كلاً منها يعني جانبًا محددًا من مشكلة الدراسة؛ مثال: أن الويب غير المرئية تستخدم لمواقع الويب الديناميكية أحيانًا، وأحيانًا أخرى لقواعد البيانات، كما أن مصطلح الويب العميقة أو deep web دائمًا ما يستخدم للتعبير عن مجموعات قواعد بيانات الويب. وقد كشف الجدول (٠-٤) أعداد استخدام كل مصطلح من المصطلحات الدالة على ظاهرة الدراسة كما يلي:

الجدول (٠- ٤) استخدام المصطلحات الدالة على الظاهرة

المصطلح	المقابل	عدد المصادر
Invisible web	الويب غير المرئية	٩
Deep web	الويب العميقة	٧

المصطلح	المقابل	عدد المصادر
Hidden web	الويب الخفية	٨
Opaque web	الويب غير الشفافة	٠
Private web	الويب ذات الملكية	٠
----	لم تذكر مسمى	٨٦
المجموع		١١٠

يدل الجدول (٠-٤) على الملاحظات التالية:

١- إن تقارب أعداد استخدام المصطلحات على الرغم من صغرها، يعني أنه ما زال هناك عدم استقرار أو دلالة قاطعة لماهية وطبيعة هذه المصطلحات. كما أن صغر أعدادها يدل على عدم رسوخ جوانب وأبعاد الظاهرة المدروسة بعد.

٢- لم تذكر أغلبية المصادر وعددها ٨٦ مسمى محددًا لهذه الظاهرة، وذلك إما للتعبير عنها بشكل مختلف؛ مثل: عنوان (استرجاع محتوى قواعد بيانات الويب)، أو التركيز على جانب محدد من المشكلة؛ مثل: برنامج الزاحف في محركات البحث.

ثانيًا: اتجاهات حل مشكلة الويب غير المرئية (إما معالجة الويب أو معالجة الأدوات):

إن معالجة مشكلة الويب غير المرئية قد أخذت الكثير من اتجاهات المعالجة التي تتنوع بين التركيز في معالجتها على أدوات البحث، أو التركيز على قواعد ومواقع الويب ذاتها. ومن ثم حاول الباحث في هذا الصدد قياس أي الاتجاهات غلب على مصادر الدراسة، وأيها لم ينل قسطًا كبيرًا من الاهتمام. وتتعدد اتجاهات معالجة الظاهرة على نواحٍ عدة؛ أهمها: التأكيد على أن دراسة قواعد ومواقع الويب غير المرئية ذاتها يمكن أن يقود إلى حل هذه المشكلة البحثية، والخروج بحلول لها نابعة من طبيعة بناء قواعد ومواقع الويب، واتجاه ثانٍ يركز

على أن الوصول لمحتوى الويب غير المرئية، يتركز على دراسات آليات عمل أدوات بحث الويب ذاتها، والوقوف على آليات العمل التي يمكن إحداث التطوير بها، حتى يمكن الوصول إلى قلب الويب غير المرئية، ولذلك فقد ضم الباحث مجموعات المصادر التي تتناول عملية الزحف وبرنامج الزاحف إلى عدد المصادر التي تخص محركات البحث. وقد غلبت فكرة دراسة الويب بقواعدها ومواقعها على فكرة تحليل أداء محركات البحث في مجملها ؛ حيث بلغ عدد مصادر معالجة الويب إجمالاً ٣٥ مصدرًا، بينما نالت أدوات البحث ١٤ مصدرًا، كان أغلبها مركزًا على محركات بحث الويب غير المرئية، بما في ذلك ملامح عمل الزاحف.

لقد وجد اتجاه ثالث يبعد قليلاً عن جانبي معالجة الويب ومعالجة أدوات البحث ؛ حيث اختص هذا الجانب بالتركيز على أهمية عوامل أخرى في حل مشكلة الويب، يمكن بمعالجتها الوصول إلى المحتوى غير المرئي. وتمثلت هذه الجوانب في التعرض لآليات البحث المتعددة والتركيز على خصائص عمل البحث المتعدد، أو التركيز على معايير البحث المتعدد مثل Z39.50، أو التركيز ملامح واجهات البحث ونماذج البحث التي توجد بها، وقد نال هذا الاتجاه الحظ الأوفر من مصادر البحث، حيث أخذ ٦١ مصدرًا من مجموع مصادر الدراسة.

الجدول (٠- ٥) اتجاهات حل مشكلة الويب غير المرئية

اتجاه معالجة المشكلة		**عدد المصادر**
معالجة الويب	المواقع	١١
	القواعد	٢٤
معالجة أدوات البحث		١٤
جوانب أخرى		٦١
المجموع		**١١٠**

يوضح الجدول (٠-٥) المعطيات التالية:

١- سيطر اتجاه البحث في خصائص الويب غير المرئية من قواعد ومواقع على فكرة دراسة آليات عمل محركات البحث، وتحليل جوانب أدائها لوضع خطة توافق بينها وبين قواعد بيانات الويب غير المرئية.

٢- إن قواعد بيانات الويب قد مثلت الجانب الأكثر أهمية في مشكلة الوصول للمحتوى غير المرئي، وذلك لما لها من طبيعة بناء خاصة لا تتوافق وبنية برنامج زاحف محركات البحث، فضلاً عن أن قواعد البيانات إنما وضعت على الويب للبحث المباشر وليس للتكشيف من جانب محركات البحث، بالإضافة إلى أنها تمتلك ملامح خاصة مثل حقوق الملكية والجوانب التجارية ورسوم تقديم الخدمات.

٣- تدل أعداد المصادر بالجدول (٠- ٥) في صغرها وتقاربها بين الويب وأدوات البحث على أن معالجة الويب غير المرئية، ما زالت في مهدها تسعى حائرة بين معالجة الويب وإعادة التفكير في بنية قواعد ومواقع الويب، وبين التفكير في دراسة آليات العمل داخل أدوات البحث، وإعادة هيكلتها مرة أخرى بما يتوافق وإمكانات بحث واسترجاع المحتوى غير المرئي.

ثالثًا: أداة البحث الأقرب للمشكلة (محركات البحث أو أدلة البحث):

إن محركات وأدلة بحث الويب المرئية دائمًا ما كانت تعمل بشكل منفصل، غير أن اتجاه الدمج بين آليتي العمل تحقق مؤخرًا متمثلاً في محرك البحث google الذي سعى إلى بناء دليل لبحث مواقع الويب يدعم أسلوب البحث الحر، وتمثل أيضًا في دليل البحث yahoo الذي دعم أسلوب البحث الحر ليجمع في عمله بين أسلوبي التصفح وإمكانات البحث الحر بالكلمات المفتاحية. ولقد لاحظ الباحث أن معظم التطبيقات الفعلية لأدوات بحث الويب غير المرئية تتجه

إلى أدلة البحث، غير أن الجانب التنظيري غالبًا ما يرد مصطلح محرك بحث الويب غير المرئية. وهو ما عكسه واقع مصادر الدراسة حيث بلغ عدد المصادر التي ذكرت محرك البحث ١٣ مصدرًا، بينما جاء مصدر واحد فقط بذكر مصطلح أدلة بحث الويب غير المرئية. وبذلك فقد بلغ عدد المصادر التي تتناول أدوات البحث وجوانبها المختلفة ٢٩ مصدرًا فقط من إجمالي مصادر الدراسة البالغ ١١٠ مصادر. وقد عمل الباحث في هذه الدراسة على تبني فكرة الجمع بين أسلوبي البحث الحر والبحث بالتصفح في بناء بوابة الويب غير المرئية.

الجدول (٠-٦) أداة البحث الأقرب للمشكلة

نوع أداة البحث	عدد المصادر
محركات البحث	١٣
أدلة البحث	١
جوانب أخرى	١٥
المجموع	**٢٩**

يوضح الجدول السابق ما يلي :

١- أنه على الرغم من الاتجاه العملي إلى بناء أدلة بحث لعرض محتوى الويب غير المرئية، إلا أن التركز الاصطلاحي والدراسي لمفهوم محرك البحث كان السائد في المصادر التي تناولت معالجة أدوات البحث. ومن ثم يمكن القول إن أدلة البحث غلبت محركات البحث في الناحية التطبيقية، بينما غلبت محركات البحث في الناحية التنظيرية لمشكلة الويب غير المرئية.

٢- عملت مصادر الدراسة على التركيز على جوانب أخرى غير محركات وأدلة البحث، تمثلت هذه الجوانب في معالجة نموذج البحث العام الذي يمكن من خلاله التفاعل مع أكثر من نموذج بحث مختلف لقواعد بيانات الويب، أيضًا التركيز على عملية دمج النتائج المسترجعة من قواعد

البيانات عند إجراء البحث بها، وكيفية فرزها وإعادة ترتيبها مرة أخرى للعرض النهائي أمام المستفيد، كما تم التركيز على جانب بنية برنامج الوسيط الذي سيكون بمقام مدير العمل داخل محرك بحث الويب غير المرئية، في استقبال استفسارات البحث وتوجيهها إلى قواعد البيانات ثم استقبال النتائج النهائية ومعالجتها.

رابعًا: التركيز على معالجة برامج الزاحف داخل محركات البحث(المشكلة في سبب الوصول أو الاسترجاع):

حاولت بعض المصادر من أدبيات الإنتاج الفكري التركيز على الدور المهم الذي يمكن أن يقوم به برنامج الزاحف في بحث واسترجاع محتوى الويب غير المرئية. كما أن الباحث وجد النسبة الغالبة من مصادر الدراسة التي تدور حول أدوات البحث وقد اهتمت بعملية الزحف والوصول إلى قواعد بيانات الويب. ويرجع هذا الاهتمام إلى طبيعة عمل برنامج الزاحف أو العنكبوت داخل محرك بحث الويب غير المرئية المقترح؛ حيث إن خطوات الوصول إلى قواعد البيانات تبدأ عند الزاحف، وإلا فلن يتمكن محرك البحث من الوصول إلى نماذج بحث وواجهات قواعد بيانات الويب، بالإضافة إلى أن التغيير المرتقب لآليات محركات بحث الويب سيطرأ أول ما يطرأ على برنامج الزاحف؛ الذي سيغير وجهته من التعامل مع صفحات الويب html pages إلى التعامل مع نماذج بحث الويب html search forms. وتبعًا لما سبق فإن الباحث حاول تحليل مصادر الدراسة لقراءة مدى اهتمامها بالتغطية الموضوعية لبرنامج الزاحف.

الجدول (٠- ٧) برنامج الزاحف في مصادر الدراسة

اتجاه تركيز المعالجة	عدد المصادر
معالجة الزاحف	٧
الجوانب الأخرى	٦
المجموع	**١٣**

يوضح الجدول (٠-٧) ما يلي:

١- أنه من بين ١٣ مصدرًا اختصت بتناول محركات بحث الويب، كان أكثر من نصفها مركزًا على تقنية الزحف وآليات عمل برنامج الزاحف المقترح في محرك بحث الويب غير المرئية.

٢- يعكس واقع مصادر الدراسة أن الجانب الأكثر أهمية وتأثيرًا في تطوير محركات بحث الويب وجعلها أكثر قدرة على الوصول إلى المحتوى غير المرئي، يتمثل في برنامج الزحف. وذلك لأن مرحلة الزحف هي المرحلة المؤثرة في عمل تنظيم واسترجاع مجموعات قواعد بيانات الويب.

خامسًا: التوزيع الزمني لمصادر الدراسة:

يعكس التوزيع الزمني لمصادر الدراسة طبيعة وبداية تناول الظاهرة، إلى جانب فترات النشاط والخمول التي أصابت أدبيات الإنتاج الفكري لهذه الظاهرة. وقد وجد الباحث أن البداية الحقيقية لهذه الظاهرة بدأت بعد عام ٢٠٠١م، وذلك لأن اهتمام الإنتاج الفكري عن المعلومات الرقمية والويب، كان منصبًا على معالجة أدوات بحث الويب المرئية، وطبيعة المعلومات النصية والمصورة والصوتية المنشورة على الويب. كما يعكس واقع المصادر أيضًا أن الاهتمام بالويب غير المرئية واسترجاع محتوى قواعد البيانات في تزايد مستمر، فضلاً عن أن هذه الظاهرة لم تجد الحل الكافي أو تصدر مجموعة من الدراسات التي تعطي لنا الحل النهائي للتغلب على مشكلة الويب غير المرئية.

يرى الباحث أن المصادر التي نشرت قبل عام ٢٠٠٢م لا ترتبط ارتباطًا وثيقًا بظاهرة الويب غير المرئية، وذلك لأن مختلف المصادر التي نشرت في الفترة بين ١٩٩٧-٢٠٠١ تتناول جوانب أخرى خلاف أدوات بحث واسترجاع الويب غير المرئية، مثل: الحديث عن بنية قواعد البيانات، ونظم إدارتها، ولغات البرمجة التي تعتمد عليها في بناء ملفاتها وجداولها.

الجدول (٠-٨) التوزيع الزمني لمصادر الدراسة

سنة النشر	عدد المصادر
١٩٩٧م	٤
١٩٩٨م	٢
١٩٩٩م	١
٢٠٠٠م	٣
٢٠٠١م	٨
٢٠٠٢م	١٠
٢٠٠٣م	١٠
٢٠٠٤م	١٠
٢٠٠٥م	١٨
٢٠٠٦م	١٦
٢٠٠٧م	١٩
٢٠٠٨م	٦
بدون تاريخ	١٣
المجموع	**١١٠**

سادسًا: التوزيع الموضوعي لمصادر الدراسة:

تتفرق ظاهرة الويب غير المرئية بين الكثير من الموضوعات الفرعية التي تصب جميعًا في مصب تنظيم واسترجاع المحتوى غير المرئي. وقد غلب هذا الواقع على مصادر الدراسة ؛ حيث انقسمت مصادر الدراسة إلى الكثير من الموضوعات الفرعية التي تتكامل معًا لأجل حل مشكلة الويب غير المرئية. وهذا إنما يأتي توافقًا مع تعدد الجوانب التي تحتاج إلى تطوير وتحديث لمعالجة هذه الظاهرة. فقضية الويب غير المرئية تحتاج إلى دراسة جانب محركات البحث وخصائص عمل الزاحف، أيضًا تحتاج إلى دراسة طبيعة وبنية قواعد بيانات الويب، كما

تحتاج إلى الوقوف على ماهية تقنيات التحليل الموضوعي الآلي من تكشيف وتصنيف آلي للتسجيلات الببليوجرافية ... وغير ذلك من الجوانب الأخرى.

الجدول (٠-٩) التوزيع الموضوعي لمصادر الدراسة

الموضوع	عدد المصادر
ماهية المعلومات والويب غير المرئية	٢٠
الزاحف	٧
قواعد البيانات	٢٤
أدوات البحث (محركات وأدلة)	١٤
تقنيات التعامل مع الويب غير المرئية	٢٥
قضايا مرتبطة	٢٠
المجموع	١١٠

يوضح الجدول (٠-٩) ما يلي:

١- أن مختلف الجوانب الموضوعية لظاهرة الويب غير المرئية تقف تقريبًا على قدر متساوٍ من الأهمية الموضوعية، وهو ما يعكسه تقارب عدد المصادر الخاصة بكل موضوع فرعي.

٢- لقد تم تناول الكثير من التقنيات التي لها دور مهم في تفعيل استرجاع محتوى الويب غير المرئية ؛ مثل بناء واجهات البحث ونماذج البحث، والتعرض لبرامج التكشيف والتصنيف الآلي لمحتوى قواعد البيانات.

٣- لم تنل أدوات البحث القدر الكافي من دراسات الويب غير المرئية، وهذا إنما يأتي تحيزًا لأهمية هذا الجانب وتفعيله في حل مشكلة استرجاع الويب غير المرئية. غير أنه كان من الواقعي التركيز على عمل برنامج الزاحف، وهذا لأن برنامج الزاحف الخطوة الأولى في عمل محرك البحث عامة، فإذا ما صلحت صلح عمل مختلف البرامج الأخرى، والعكس من ذلك صحيح.

سابعًا: التوزيع الجغرافي واللغوي لمصادر الدراسة:

لم يتطرق الإنتاج الفكري العربي إلى الحديث عن ظاهرة الويب غير المرئية، أو عن بحث واسترجاع مجموعات قواعد بيانات الويب باستخدام محركات وأدلة البحث. وما استعان به الباحث من مصادر عربية إنما لتناول نقاط فرعية غير متخصصة أو الوقوف عليها في تحديد ملامح بحث واسترجاع المحتوى غير المرئي. وقد عكس ذلك الفارق الكبير في نسبة المصادر العربية إلى نسبة المصادر الأجنبية كما يتضح من الجدول (٠-١٠).

الجدول (٠-١٠) التوزيع الجغرافي اللغوي لمصادر الدراسة

اللغة والثقافة	عدد المصادر
الأدبيات العربية	٥
الأدبيات الأجنبية	١٠٥
المجموع	١١٠

وستتكون الدراسة من الفصول الآتية :

الفصل الأول : الويب : البنية والتنظيم والاسترجاع.

الفصل الثاني : قواعد البيانات " الويب غير المرئية".

الفصل الثالث : أدوات تنظيم وبحث الويب غير المرئية.

الفصل الرابع : واقع قواعد بيانات الويب غير المرئية " الدراسة التجريبية ".

الفصل الخامس : محركات وأدلة الويب غير المرئية " الدراسة التجريبية ".

الفصل السادس : مواصفات محركات بحث قواعد بيانات الويب غير المرئية.

وذيلت الدراسة بمجموعة من النتائج والتوصيات وقائمة بالمصادر والمراجع التي اعتمدت عليها الدراسة، بالإضافة إلى قائمة بأهم المصطلحات مرتبة هجائيًا.

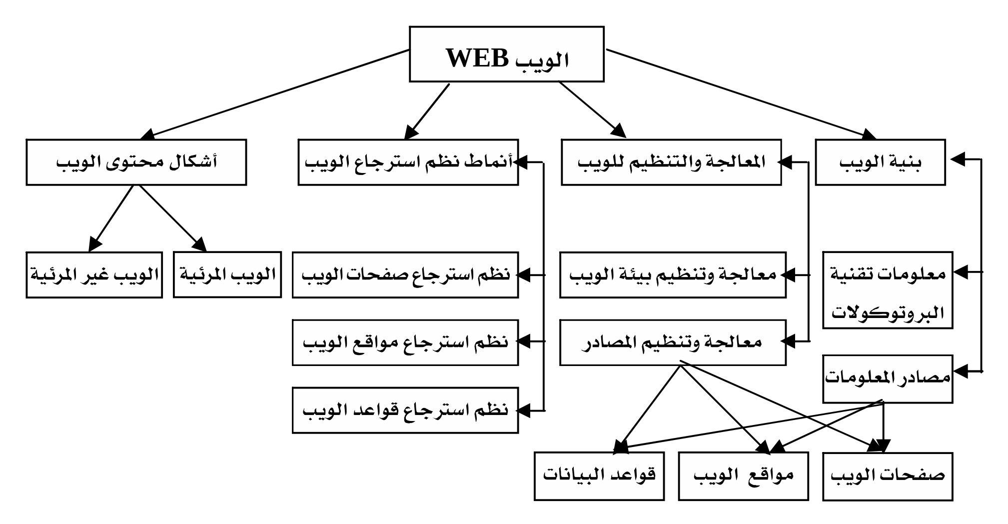
الفصل الأول

الويب ..

البنية والتنظيم والاسترجاع

١ / ٠ تمهيد:

كونت الويب عالمًا من المعلومات الرقمية متعدد الموضوعات مختلف الأشكال ويختلف هذا العالم الرقمي عن نظيره التقليدي في التعامل مع المعلومات معالجة وتنظيمًا واسترجاعًا. تطورت أشكال مصادر معلومات الويب منذ بدايتها لتمر بالأشكال النصية والسمعية والمرئية، وصولاً إلى ما يعرف بمراصد أو قواعد البيانات. وانقسمت الويب منذ ظهور قواعد البيانات إلى شقين؛ الأول: مرئي متمثلٌ في صفحات ومواقع الويب، والثاني: غير مرئي أو عميق أو مخفي متمثل في ملفات قواعد البيانات. وتتناول الدراسة بداية من هذا الفصل دراسة ماهية الويب عامة والويب غير المرئية خاصة، ثم تتابع الفصول للتعرف إلى خصائص بيئة الويب غير المرئية وسبل التعامل الفني من تنظيم واسترجاع بما يتناسب مع طبيعة بناء مصادر قواعد البيانات. وتأتي أهمية الدراسة في محاولة إيجاد الشكل الملائم من تعامل محركات البحث مع قواعد البيانات إضافة وتنظيمًا وبحثًا. غير أن رؤيتنا للويب سوف تختلف عن نظرة الدراسات الأخرى للويب ؛ حيث تركز الدراسة هنا على الجوانب التنظيمية والهيكلية والعوامل التي تقف خلف ظهور الويب بما نشاهدها عليه عبر نوافذ استخدام الإنترنت. وتأتي عناصر هذا الفصل لتركز على بيئة الويب بنية وتنظيمًا ومصادرَ، وصولاً إلى تناول أحد مصادر الويب وهي قواعد البيانات التي تشكل بمفردها بيئة جديدة تسمى بيئة الويب غير المرئية. ويوضح الشكل (١-١) ملامح الويب من حيث البنية والتنظيم:



الشكل (١-١) بنية وتنظيم الويب وبيئة الويب غير المرئية

١/١ بنية الويب:

قبل التعريف ببنية الويب، يمكن وصف الويب على أنها "شبكة عالمية تقدم خدمات الإنترنت، وتجهيز الوثائق التي يعتمد بناؤها على لغة ترميز النص الفائق التي تسمح باسترجاع الوثائق باستخدام الروابط الفائقة، وتم تصميم الويب في عام ١٩٨٩م عن طريق طالب في جامعة إلينوي University of Illinois وارتبطت بداية الويب باستخدام متصفح Mosaic ثم خروج متصفح نيتسكيب Netscape Web browser " وذلك وفق ما جاء في قاموس مصطلحات علم المكتبات والمعلومات على الخط المباشر ODLIS تحت مسمى الشبكة العالمية العنكبوتية World Wide Web (WWW). ويتفق الباحث مع معنى المفهوم على أن الويب كشبكة عالمية إنما هي جزء من الإنترنت تقدم خدمات ذات طابع يتفق مع بنية وتنظيم الويب[١].

(١) ODLIS —Online dictionary for library and information science. Cited 10/5/2007, cited at http://lu.com/odlis/odlis_w.cfm

وتعتمد الويب كبيئة رقمية في بنيتها على شقي بنية النظام الرقمي ؛ وهما الجانب المادي hard ware من التجهيزات والجانب الفني من البرامج soft ware. غير أن الحديث هنا سيركز فقط على الجانب الفني من بنية الويب وهي البرامج والمعلومات والنظم التي تعمل بها الويب. وتحتوي الويب على عناصر معلوماتية تختلف بين بروتوكولات عمل الويب وتنظيم المعلومات حفظا واستدعاء، بالإضافة إلى المعلومات المسجلة والموزعة على أشكال وموضوعات مختلفة. وتم توضيحها كما يلي :

١/١/١ المعلومات الفنية (بروتوكولات الويب) :

تتنوع بروتوكولات عمل الويب نظرا لتنوع خدمات المعلومات التي تقدمها. غير أن أشهر هذه البروتوكولات هو بروتوكول نقل النص الفائق HTTP، وبروتوكول نقل معلومات الإنترنت IP، وبروتوكول البريد الإلكتروني SMTP. ويوضح الجدول (١-١) مختلف البروتوكولات التي تتعامل بها الويب[١] :

الجدول (١-١) بروتوكولات عمل الويب

Protocol	Description
ARP	Address Resolution Protocol
HTTP	Hypertext Transfer Protocol
FTP	File Transfer Protocol
SMTP	Simple Mail Transfer Protocol
DNS	Domain Name Service
DSN	Data Source Name
IP	Internet Protocol
TELNET	TCP/IP Terminal Emulation Protocol

(١) Web protocols. securepaynet, 2006, cited 10/5/2006, Cited at http://www.realifewebdesigns.com/web-resources/web-protocols.html

Protocol	Description
IMAP	Internet Message Access Protocol
ICMP	Internet Control Message Protocol
IDRP	ICMP Router-Discovery Protocol
DHCP	Dynamic Host Configuration Protocol
IRC	Internet Relay Chat Protocol
POP3	Post Office Protocol version 3
PAR	Positive Acknowledgment and Retransmission
RLOGIN	Remote Login
SSL	Secure Sockets Layer
SSH	Secure Shell
TCP	Transmission Control Protocol
UPD	User Data gram Protocol
UPS	Uninterruptible Power Supply

يعد بروتوكول http أهم البروتوكولات السابقة لما له من ارتباط وثيق بآلية عمل الويب؛ حيث يعد التصفح والإبحار داخل الويب web navigation أهم أساليب استرجاع المعلومات على الويب. يلي ذلك بروتوكولات خدمات الويب الأخرى مثل البريد الإلكتروني SMTP ونقل الملفات FTP.

٢/١/١ مصادر معلومات الويب:

يقصد بمصادر معلومات الويب هنا مختلف أشكال حفظ المعلومات المراد إتاحتها عبر الويب بما تحويه من وسائط مختلفة لحمل المعلومات مثل النص والصوت والصورة. وقد اقترح الباحث تقسيم مصادر المعلومات بحسب إمكانات حفظ وعرض المعلومات بدلاً من التقسيم بحسب أشكال المعلومات ذاتها ، وذلك اتساقًا مع الحديث عن تنظيم المعلومات داخل بنية الويب.

وقد استثنى الباحث من هذه المصادر الحديث عن الملفات الرقمية التي تنتمي إلى صفحات ومواقع الويب، لما لهذه الملفات من خصائص تختلف بها عن منظومة بنية الويب. وفيما يلي إيضاح لجوانب هذا الاختلاف:

- تعتبر الملفات الرقمية وحدات لا يمكن تقسيمها عكس صفحة الويب التي تعد وحدة بناء للويب، وتحتوي هذه الوحدة على ملفات تحمل معلومات متعددة الوسائط.
- إن ملفات الويب لا يمكن تواجدها على الويب بمفردها دون صفحة مضيفة host web page لهذه الملفات تحويها في إطار عرض الموضوع العام.
- إن ملفات الويب لا تحمل خصائص شكلية تؤثر في تنظيم الويب، وإنما تحمل الملفات الرقمية خصائص ما تحمله من معلومات مثل خصائص النص للملفات النصية أو خصائص المعلومات المرئية لملفات الصور الثابتة والمتحركة أو خصائص المعلومات السمعية لملفات الصوت.
- تحتاج الملفات الرقمية إلى المعالجة الفنية بمعايير معالجة وسائط حمل المعلومات "النص والصورة والصوت"، بينما تحتاج صفحات ومواقع وقواعد البيانات إلى معايير معالجة مصادر المعلومات المكونة من خصائص شكلية وموضوعية.

١/٢/١/١ صفحات الويب:

عرف قاموس مصطلحات المكتبات والمعلومات ODLIS صفحة الويب على أنها " وثيقة إلكترونية تمت كتابتها بلغة ترميز النص الفائق HTML ويتم استرجاعها باستخدام متصفح الويب عن طريق محدد المصادر الموحد URL، وتمثل صفحة الويب في أحيان كثيرة تجميع لمجموعة روابط ملفات إلكترونية ذات أنواع مختلفة، ويمكن لصفحة الويب أيضا احتواء أشكال مختلفة لوسائط حمل المعلومات مثل النص والصوت والصورة[١]."

---

(١) ODLIS —Online dictionary for library and information science. OP.CIT.

وتتكون بنية صفحات الويب من أجزاء أساسية يحتوي كل منها على معلومات ذات هدف مختلف عن الآخر. حيث تبدأ صفحة الويب **أولاً** برأس الصفحة page head الذي يتم التعبير عنه بـ<head> content </head> بما يحدد بداية المعلومات ونهايتها. ويخصص رأس الصفحة لتحديد موضوع الصفحة ووصف الكلمات الدالة بها ، ويساعد هذا الجزء محركات البحث في التعرف إلى محتوى صفحة الويب. **ثانيًا:** جسم صفحة الويب: ويحتوي على المعلومات المراد بثها لمستخدمي الويب بما في ذلك من وسائط مختلفة لحمل المعلومات. ويحدد هذا الجزءَ عنصرا الترميز <body> content </body>. ثم خاتمة صفحة الويب التي تحوي حقوق الملكية الفكرية وحقوق النشر.

وهناك ثمة جوانب لابد من مراعاتها عن بناء صفحات الويب صاغها اتحاد شركات الويب W3C ، وهي كالتالي(١):

- أن يتسق تصميم صفحات الويب مع متطلبات مستخدمي صفحة الويب: حيث يراعي التصميم احتياجات المرحلة العمرية والجنس والحالة المعيشية لمستخدمي الصفحات المتوقعين.
- سهولة الإبحار والتصفح داخل محتويات الصفحات والصفحات الداخلية المكونة لمواقع الويب.
- التأكيد على سلامة الروابط إلى الصفحات الأخرى داخل صفحة الويب، وعدم وجود روابط ضائعة Broken Links.
- التأكيد على جودة محتوى صفحات الويب بما يحقق كفاءة أعلى داخل محركات البحث، فضلاً عن تحقيق أعلى درجة تصفح.
- أن يتم تكرار تحديث صفحات الويب بشكل دوري على فترات زمنية قصيرة.

(١) web designs. Realife web designs, 2005, cited 15/5/2007, cited at http://www.realifewebdesigns.com/web-resources/web-tips.html

• إمكانية التمتع بتقديم نشرات إخبارية E-News letters تذكر مستخدمي صفحات الويب من حين إلى آخر بزيارة هذه الصفحات.

٢/٢/١/١ مواقع الويب:

عرف قاموس مصطلحات المكتبات والمعلومات على الخط المباشر ODLIS موقع الويب بأنه "مجموعة من صفحات الويب المتكاملة المتصلة معا بروابط فائقة ومخزنة على أحد خادمات الويب، وتستخدم من خلال أحد برامج تصفح الويب. وهذه الصفحات متاحة ٢٤ ساعة في اليوم بحيث يمكن لمستخدمي الويب الاطلاع عليها في أي وقت. ويتم بناء صفحات الموقع نتيجة لعمل مؤسسة أو هيئة أو فرد على اختلاف اتجاهاتهم، ويطلق على الشاشة الأولى للموقع اسم الصفحة الدليلية Home page. ويجب على مصمم موقع الويب متابعة صفحاته لتحديثها من حين إلى آخر." ويتميز مفهوم قاموس مصطلحات المكتبات والمعلومات على الخط المباشر في جملته بالشمولية؛ حيث تطرق إلى جوانب الإتاحة والتحديث والاسترجاع من خلال متصفحات الويب[١].

إن مواقع الويب، وان اختلفت في أهدافها وعدد صفحاتها، إلا أنها تسعى إلى اجتذاب اكبر عدد ممكن من مستخدمي الويب، ولذلك فإن على مصممي مواقع الويب الاهتمام بمجموعة من خصائص مصادر المعلومات التي توفر للموقع أكبر عدد من الزيارات وأعلى درجة من الأداء، وهذه الخصائص حددها Alastair Smith في حديثه عن تقييم مواقع الويب، وهي كما يلي[٢]:

أولاً: مجال التغطية لموقع الويب: حيث يجب على الموقع توضيح حدود الموضوع داخله مثل:

(١) ODLIS —Online dictionary for library and information science. OP.CIT.

(٢) Smith, Alastair G. Testing the surf: criteria for evaluating internet information resources." The Public-Access computer systems review 8, no. 3 (1997). Cited 20/5/2007, cited at http://info.lib.uh.edu/pacsrev.html.

❖ مدى الشمولية في عرض المعلومات.

❖ مدى العمق أو التخصص في الموضوع.

❖ تحديد الفترة الزمنية التي تعالج فيها معلومات الموقع.

❖ طبيعة أشكال مواد المعلومات التي يحتوي عليها الموقع.

ثانيًا : محتوى موقع الويب كمصدر معلومات: حيث يجيب موقع الويب على الأسئلة التالية:

- هل يحمل الموقع معلومات أولية أم متخصصة في الموضوع؟
- هل يحمل الموقع المعلومات ذاتها أم روابط وإشارات ببليوجرافية فقط عن الموضوع؟
- هل يتكامل الموقع مع مواقع أخرى متخصصة أم يقف بشكل فردي؟
- ما طبيعة المعلومات التي يحملها الموقع ، نصية أم صوتية أو مرئية؟

ويشمل الحديث عن محتوى موقع الويب الجوانب التالية:

١- **دقة المعلومات:** هل يتبع هذا الموقع هيئة علمية متخصصة في الموضوع أم يعتمد على باقي مواقع الويب في استقاء المعلومات المعروضة داخله؟

٢- **المسئولية الفكرية**: فلمن تتبع هذه المعلومات؟

٣- **الحداثة** : وتتعلق بطبيعة موقع الويب في ثبات المعلومات أم إجراء التحديث الدوري لمحتوى صفحاته.

٤- **التفرد في المحتوى:** ويتعلق ذلك باختلاف الموقع عن غيره من مواقع الويب في عرض معلومات الموضوع، أو أن يختلف موقع الويب في وسائط حمل المعلومات مع اتفاقه في المحتوى مع المواقع الأخرى.

٥- **الروابط مع المواقع الأخرى:** حيث يوفر الموقع الوصول إلى مواقع الموضوع الأخرى، فضلاً عن تحديث روابط صفحاته ذاتها وصيانتها من حين إلى آخر، أيضًا مدى حداثة مصادر المعلومات التي يرشد إليها هذا الموقع وتدل عليها روابطه إلى الصفحات الأخرى.

٦- **دقة التنسيق والصياغة:** حيث تهتم مواقع الويب بتنسيق المتن داخل صفحاتها، وفي حالة التحول إلى وسائط أخرى مثل المصادر المرئية فيجب على الموقع الاهتمام بمدى وضوح المصادر المصورة وكثافتها.

ثالثًا: الرسومات والوسائط المتعددة:

يرتبط هذا العنصر بمدى قابلية المشاهدة والاطلاع لصفحات الموقع، إضافة إلى ملاءمة الأشكال المصورة لمحتوى الموقع وتوافقها مع موضوع موقع الويب.

رابعًا: الغاية والجمهور :

ويقصد بذلك أهمية التخطيط المسبق لبناء موقع الويب والتعريف بالهدف من تصميمه، بالإضافة إلى مراعاة التوافق بين حاجات مستخدمي الويب وبين هدف ومقصد إنشاء موقع الويب.

خامسًا: تقييم المراجعات العلمية:

حيث يؤثر في تقييم مواقع الويب مدى الاستشهاد به وتقييمه في المراجعات العلمية، بالإضافة إلى تعليقات الهيئات العلمية والأفراد على مدى الإفادة من محتوى موقع الويب.

سادسًا: كفاءة الأداء :

يهتم مستخدمو الويب بمدى الإفادة من استخدام مصادر الويب الرقمية، لأن المستفيد يسعى دائمًا للوصول إلى القدر الأكبر من الكفاءة والدقة التي قد تحققهما مواقع الويب. وتقاس كفاءة أداء الموقع لدى مستخدمي الويب من خلال مجموعة من العوامل هي:

◆ **قابلية الاستخدام وسهولته:**

يهتم أغلب مستخدمي الويب بمدى سهولة تصفح مواقع الويب والبحث داخلها، إلى جانب ذلك مدى وضوح عناصر تشغيل الموقع والحصول على واجهة البحث الواضحة وأوامر التشغيل سهلة الاستيعاب.

◆ **متطلبات التشغيل والتصفح:**

يتساءل مستخدمو الويب دائمًا عن متطلبات تشغيل مواقع الويب من حيث احتياج هذه المواقع إلى برامج تشغيل خاصة أو متطلبات أخرى لشبكة الويب أو الاشتراك لدى الموقع أو الاحتياج لكلمات المرور، بالإضافة إلى نظم التشغيل المتوافقة مع موقع الويب أو خادم الويب.

◆ **إمكانات بحث المعلومات:**

تمثل قدرات البحث والاسترجاع أهم جوانب بناء مواقع الويب ذات الكم الكبير من المعلومات. ومن ثم فإن هناك مجموعة من التساؤلات التي يطرحها المستفيد على نفسه عند استخدام مواقع الويب مثل : ما إمكانات البحث التي يوفرها مصدر المعلومات؟ وما معاملات وآليات البحث داخل موقع الويب؟

◆ **تنظيم المعلومات وتصفحها:**

يمكن لموقع الويب توفير تقسيمات موضوعية تسهل لمستخدمي الموقع التحرك بين الموضوعات وتفريعاتها، واسترجاع المعلومات المخصصة والدقيقة المرتبطة بمصطلحات البحث. فضلاً عن اعتماد موقع الويب لتنظيم صفحاته بما يسهل للمستخدم تصفح محتوياته بسهولة ويسر.

◆ **مدى التفاعلية:**

تتصف مواقع الويب بالتفاعلية عند تقديمها شاشات المساعدة ورسائل الحوار إلى مستخدمي الموقع عند الحاجة إليها؛ مما يمكن المستخدم من فهم آليات عمل

موقع الويب والحصول على ما يريده من معلومات. على الجانب الآخر فإن التفاعلية تعني لبعض مستخدمي الويب مدى قدرة الموقع على تطوير وتحديث محتواه الموضوعي على فترات قصيرة، بما يكفل لمستخدمي الويب الحصول على أحدث المعلومات في التخصص.

- **الوصول والإتاحة:**

يحتاج مستخدمو الويب إلى سهولة الاتصال بموقع الويب، وإمكانات تحميل محتوى الموقع بالسرعة المطلوبة، حيث تتصف بعض المواقع ببطء الوصول إليها وبطء تحميل محتواها على نوافذ الويب. إلى جانب عنصر الإتاحة المتمثل في الحصول على المعلومات بدون اشتراك أو دفع مقابل للمعلومات أو احتياج كلمات المرور.

سابعًا: التكلفة:

إن عامل التكلفة هو أهم عوامل الحكم على كفاءة موقع الويب. حيث تقاس التكلفة من جانب مستخدمي الويب بعنصرين هما:

- تكلفة الاتصال بموقع الويب.
- تكلفة الحصول على المعلومات المتاحة داخل صفحات موقع الويب.

وتحتاج مواقع الويب إلى جانب معايير البناء والتصميم السابقة إلى عمليات المتابعة والرعاية الشاملة. وتتمثل جوانب رعاية مصادر الويب في العناصر التالية[1] :

١- متابعة التطوير والتحديث الدائم لموقع الويب بما يتماشى مع ما يستجد من برامج تشغيل المصادر الرقمية، إلى جانب التوافق مع بيئات التشغيل المتطورة. فضلاً عن توفير عناصر أمن وحماية المعلومات ضد اختراقات مستخدمي الويب.

(١) Preservation risk management for web resources. D-Lib magazine, 2002, cited at 25/5/2007, cited at http://www.dlib.org/dlib/january02/kenney/01kenney.html

٢- متابعة الإجراءات الإدارية مع خادم الويب الذي يحتفظ بالموقع داخل قاعدة بياناته، وما يتبع ذلك من تحديث وتطوير لمسار موقع الويب domain name.

٣- متابعة وصيانة موقع الويب على الشبكة بما يحفظ مصدر الويب من التوقف نتيجة للضغط أو إساءة الاستخدام.

٤- متابعة سياسات النسخ والحفظ لمصدر المعلومات الرقمي داخل خادم الويب من حيث وسيط حمل النسخ الاحتياطية للموقع وأماكن تخزين هذه النسخ.

٥- الاهتمام بتوفير موقع جغرافي مناسب لخادم الويب الذي يحوي موقع الويب ومصادر المعلومات. وذلك بما يكفل لخادم الويب البعد عن الأخطار الطبيعية مثل درجات الحرارة والرطوبة والزلازل.

وتقسم مواقع الويب من حيث الاهتمام بالتحديث والتطوير لمحتوى صفحاتها إلى قسمين هما :

القسم الأول : مواقع ثابتة:

تتصف مواقع الويب بالثبات إذا ظل محتوى صفحات الموقع دون تحديث أو تغيير. أما التفاعلية فهي تغيير محتوى صفحات الويب بما يتوافق مع مستجدات التخصص الموضوعي. غير أن صفة التفاعلية يلحقها بعض مستخدمي الويب بالمصادر التي توفر آليات المساعدة وإتاحة عناصر تشغيل الموقع أمام مستخدمي الويب بما يكفل إمكانية استرجاع معلومات الموقع بيسر تام وفي أقل وقت.

القسم الثاني : مواقع تفاعلية ديناميكية:

تحتاج بعض مواقع الويب دون بعضها الآخر لعنصر التفاعلية مع مستخدمي الويب. وتؤثر طبيعة محتوى الموقع في مدى احتياجه للتفاعلية أو ثبات المحتوى ؛ حيث تحتاج المواقع الإخبارية والإعلامية إلى تغيير مستمر لمحتوى الموقع يومًا بعد

يوم وذلك لما لطبيعة الأخبار اليومية من اختلاف. كما أن الدوريات العلمية التي تصدر على فترات قصيرة تحتاج إلى تغيير أعدادها في كل إصدارة جديدة. على الجانب الآخر فإن نوع موقع الويب يؤثر أيضًا في احتياجه للتفاعل مع مستخدمي الويب ؛ حيث لا يمكن لمحركات البحث أو البوابات تنظيم المعلومات وبحثها دون توجيه ومساعدة مستخدمي الموقع بواسطة رسائل الحوار أو شاشات مساعدة البحث والبحث المتقدم، فضلاً عن كتابة التعليقات والتغذية المرتدة.

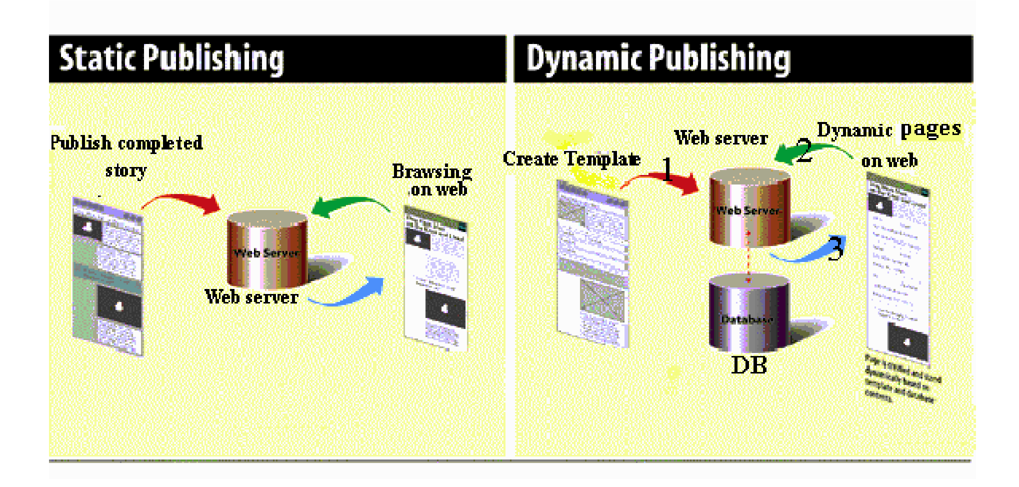
ويقدم Robert Neches نموذجًا للمواقع الديناميكية في تغيير وتحديث محتواها حيث يهتم هذا الموقع بمتابعة الأخبار الجغرافية على مستوى العام بالإضافة إلى أهم الأحداث المرتبطة بأماكن محددة وذات طابع ساخن، ومهمة هذا الموقع هو رصد الأخبار اليومية داخل مواقع الأخبار على الويب ، ثم تجميعها وإعادة تنظيمها مرة أخرى بما يتوافق ومعالجة الأماكن الجغرافية على مستوى العالم. ومن ثم يقدم موقع الويب المعلومات فور ظهورها بما يحفظ له الحداثة الدائمة. وقد حدد النموذج مجموعة من المتطلبات للحفاظ على فاعلية وديناميكية مصدر الويب. وهي[١]:

- **الشكل والاختيار:** حيث يتأثر محتوى موقع الويب بما يختاره من معلومات ومدى حداثة هذه المعلومات. إضافة إلى جودة خدمات المعلومات المقدمة.
- **إعادة تنظيم المعلومات** : يقدم موقع الويب التفاعلي مجموعة من المعلومات تم إصدارها بشكل منفصل، ويأتي دوره في مدى إعادة تنظيم وتشكيل هذه المعلومات بصورة مختلفة مرة أخرى.
- **تشكيل خدمات المعلومات** داخل مصدر الويب بما يوافق التغيير المستمر والتطوير في عرض المعلومات.

(١) Ke-Thia, Yaoand and Robert, Neches. Dynamic coordination of information management services for processing dynamic web content, Information Sciences Institute, University of Southern California, 2005, cited at 25/5/2007, cited at http://www2002.org/CDROM/refereed/613.

- **سلامة السياق العام** لعرض المعلومات: حيث قد يؤدي جمع معلومات مختلفة في إطار واحد إلى ضعف سياق عرض المحتوى العام.

ويوضح الشكل **(**١-٢**)** مواقع الويب ؛ التفاعلية والثابتة اعتماداً على التطوير والتحديث في محتوى صفحات المواقع[١]:



الشكل (١-٢) مواقع الويب التفاعلية والثابتة

يوضح الجانب الأيمن من الشكل (١-٢) صفة التفاعلية في تغيير وتحديث مصدر الويب ؛ حيث يدل السهم رقم (١) في الجانب الأيمن على نموذج النشر القابل للتغيير ، أما السهم (٢) فيدل على خادم الويب الذي يستقبل آخر التغييرات في محتوى موقع الويب، ويدل السهم (٣) على العمل المتواصل في قاعدة البيانات بإضافة وعرض أحدث تطوير للمحتوى أمام متصفحي الموقع . والجانب الأيسر من الشكل يدل على إدخال معلومات ذات شكل ثابت إلى خادم الويب ثم عرض محتوى الموقع دون الوصول إلى تغييره أو تطويره مرة أخرى.

(١) Timeless principles of design (Web Techniques, CMP technology,2007,cited at 25/7/2007, cited at http://www.ddj.com/dept/architect/184414356?pgno=1

٣/٢/١/١ قواعد البيانات:

تتميز قواعد البيانات عن باقي مصادر الويب بخصوصية في التنظيم والبناء ومعالجة الملفات الرقمية داخلها. كما أن مصادر الويب تعتمد على العرض المباشر لما تحويه من معلومات خلاف قاعدة البيانات التي تتطلب استخدام واجهة بحث ذات استراتيجيات خاصة قادرة على استدعاء الملفات الرقمية داخل قاعدة البيانات. وقد عرف قاموس مصطلحات المكتبات والمعلومات ODLIS قاعدة البيانات على أنها " مجموعة كبيرة من الملفات تحوي معلومات رقمية في أشكال مختلفة مثل البيانات البيليوجرافية والمستخلصات والصور والنصوص الكاملة... وغيرها. ينتمي جميعها لمجال موضوعي محدد وترتكز في بنيتها على مجموعة التسجيلات المعيارية بحسب طبيعة قاعدة البيانات، وذلك حتى يتم تخزينها وتنظيمها واستدعاؤها من خلال نظم إدارة قواعد البيانات DBMS. وتتنوع قواعد البيانات داخل مجال المكتبات بين قواعد البيانات البيليوجرافية كالفهارس، وقواعد بيانات المستخلصات ، وقواعد بيانات النصوص الكاملة[١]."

ويكشف المفهوم السابق لمصطلح قواعد البيانات عن أهمية برامج إدارة قواعد البيانات ومدى تأثيرها في عزل قواعد بيانات الويب غير المرئية عن محركات البحث، مما يخفيها عن أعين مستخدمي الويب. ويبرز هنا مدى الحاجة إلى التوافق بين آليات عمل نظم إدارة قواعد البيانات وبين آليات عمل محركات البحث حتى يتمكن محرك البحث في مرحلة لاحقة من التعمق داخل محتوى قواعد البيانات واستدعاء ملفاتها ضمن نتائجه أمام مستخدمي الويب. ومن بين أشهر أنواع نظم إدارة قواعد البيانات نظاما Microsoft's Access and Borland's Paradox وهما شائعا الاستخدام في أجهزة الحاسب الشخصية PCs، أيضا هناك نظم إدارة قواعد البيانات ذات الحجم والتعقيد العالي مثل أنظمة Oracle 8 or Sybase SQL or SAP.

(١) ODLIS —Online dictionary for library and information science. OP.Cit

ويساهم العنصر البشري في محاولة رفع كفاءة الأداء لقاعدة البيانات فيما يطلق عليه Database Administrators، وتحدد وظائف مديري قواعد البيانات على النحو التالي(١):

١) صيانة وتحديث نظم إدارة قاعدة البيانات.

٢) بناء وتقييم سياسات عمل قاعدة البيانات.

٣) توظيف وتدريب القائمين على عمل قاعدة البيانات.

٤) الاستعداد التام لتغيير نظم العمل وتطوير التكنولوجيا المستخدمة.

٥) متابعة تقارير عمل قاعدة البيانات.

٦) تحليل كفاءة عمل قاعدة البيانات وجودة المعلومات المستخرجة.

٧) تحديد متطلبات المستفيدين الجدد وتنمية كفاءة قاعدة البيانات.

لقد عملت قواعد البيانات منذ بدايتها على توفير ميزات افتقدتها أشكال التخزين التقليدية؛ مثل توفير سرعة الحصول على المعلومات، وتخزين أكبر قدر من البيانات، إضافة إلى عمليات التحديث والبحث عن المعلومات. وعلى ذلك يمكن صياغة مهام قاعدة البيانات في الحفظ والتنظيم بجانب سرعة البحث ودرجة الكفاءة. وجاءت الويب لتحمل محتوى مصادر قواعد البيانات إلى مستخدمي الويب بواسطة الاتصال بالإنترنت. ومن ثم تعد قاعدة البيانات أحد مصادر المعلومات الرقمية المنتمية للويب، والتي عملت على إحلال نظام تكشيف الربط اللاحق بدلاً من تكشيف الربط المسبق حتى يتمكن المستخدم من استرجاع تسجيلات قاعدة البيانات بواسطة كلمات مفتاحية داخل مختلف حقول التسجيلة. غير أن هناك من رأى أن لهذا الأسلوب في التكشيف بعض العيوب مثل(٢):

---

(١) Database Administrator jobs and information. UniXL , 2004, cited 2/6/2007, cited at http://www.unixl.com/dir/information_technology/databases/

(٢) Introduction to databases for web developers: What is a database. Extropia, 2007, cited 25/5/2007, cited at http://www.extropia.com/tutorials/sql/what_is_database.html

- ارتفاع نسبة الاستدعاء حيث يتم استرجاع تسجيلات كثيرة مشتركة فقط في بعض الكلمات الدالة.
- تحتاج بعض نظم إدارة قواعد البيانات إلى درجة التحكم العالي ودقة تغيير وبحث البيانات.
- ارتفاع تكلفة بعض برامج تشغيل قواعد البيانات.

غير أن الباحث يرى أن هذه العيوب لم تعد بالشكل المذكور سابقًا؛ لما شهدته نظم إدارة قواعد البيانات من تطوير. على الجانب الآخر فإن قواعد بيانات الويب تقسم إلى نوعين ، هما[1]:

أولاً: قواعد البيانات التحليلية: OLAP- On Line Analytical Processing

وهي قواعد البيانات التي يتم إنشاؤها لاختزان معلومات محددة والاستعانة بها بحسب الحاجة إليها، غير أن الدور الرئيس لقاعدة البيانات يتمثل في عمليات التحليل وإنتاج معلومات جديدة تدل على طبيعة واتجاهات البيانات السابق تخزينها. ومن أمثلة قواعد البيانات التحليلية على الويب قاعدة بيانات Amazon.com حيث يعمل برنامج إدارة قاعدة البيانات على تحليل أسئلة بحث مستخدمي الويب ثم إنتاج الواجهة المحددة لنتائج البحث.

ثانيًا: قواعد معالجة العمليات : OLTP On Line Transaction Processing

حيث يقوم هذا النوع بمهام قاعدة البيانات من تخزين واسترجاع للبيانات، إلى جانب تقارير حالة المعلومات من حذف وتعديل وإضافة.

وبجانب النوعين السابقين فإن قواعد البيانات تعتمد في بنائها على ثلاثة نماذج، يعتمد أحدها بحسب طبيعة عمل قاعدة البيانات والهيئة المنشئة، وهي قواعد البيانات العلائقية والهرمية والشبكية.

---

(١) Introduction to databases for web developers. Ibid.

٢/١ المعالجة والتنظيم للويب:

يشمل الحديث عن المعالجة الفنية والتنظيم لمعلومات الويب شقين أساسيين هما ؛ المعالجة الفنية للمعلومات في بيئة الويب ، والمعالجة والتنظيم لمصادر الويب الرقمية. وهي كالتالي:

١/٢/١ المعالجة والتنظيم لبيئة الويب:

افتقدت المعلومات الرقمية على الويب إلى شكل المعالجة والتنظيم الكامل داخل مصادر الويب، وقد حظيت هذه المشكلة بالدراسة من باحثي علوم الحاسبات والمعلومات. واتفق الباحثون على وضع تدرج منطقي لأسباب هذه الظاهرة يتضح في العرض التالي(١):

- أن المعلومات الرقمية على الويب دائمة التغيير بتغير الوقت.
- تحتاج المعلومات المتشابهة في الموضوع مختلفة المصادر إلى الترابط فيما بينها.
- يحتاج ربط المعلومات المتشابهة معًا إلى نظام تحليل دقيق لعملية تنظيم هذه المعلومات.
- يحتاج نظام التحليل والتنظيم الدقيق إلى رسم واضح لمهامه وإجراءاته في تنظيم المعلومات على الويب.
- إن إجراءات ومهام تنظيم المعلومات دائمًا ما تتصف بالثبات في الوقت الذي تحتاج فيه إلى التطوير الدائم.

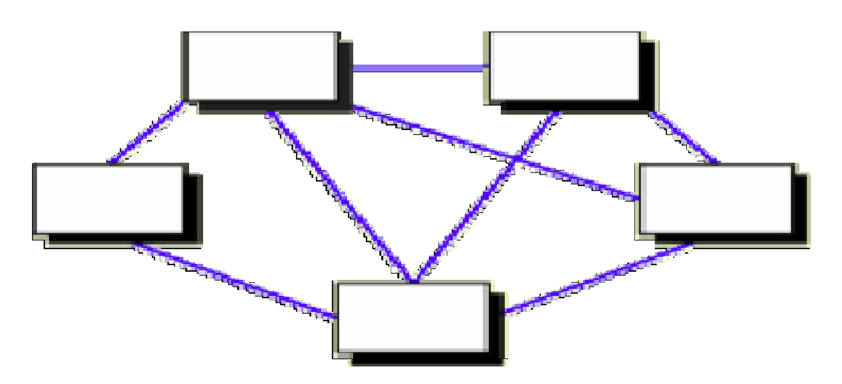
وتتضمن الويب نماذج عامة ذات أشكال متعددة لحمل وتنظيم المعلومات، وهذه الأشكال يمكن التعرف إليها كالتالي(٢):

(١) Ke-Thia Yao, Robert Neches. Op.cit.

(٢) Watson, James. Information models for web structure, published by NIVA Inc. cited 20/5/2007, Cited at http://www.writersblock.ca/spring1998/feature.htm

١/١/٢/١ نموذج الويب The Web Model

تبنى المعلومات داخل هذا النموذج باستخدام كلمات ذات دلالة في كل صفحة ويتم ربط الصفحات فيما بينها عن طريق روابط بين هذه الكلمات، والانتقال من صفحة إلى أخرى بالضغط بشكل عشوائي يأتي من متابعة كلمات محددة، ولا يرتبط بتتابع الصفحات أو التحرك داخل مستوى واحد لبناء موقع الويب. تستخدم الويب لغة خاصة تسمى HTML تأتي من مصدر عام للغات ترميز النصوص الفائقة هي لغة الترميز المعيارية العامة SGML. وتتكون الوثيقة الرقمية في هذه الحالة من عناصر يعبر عنها بحقول أو تيجان الصفحة HTML Tags.

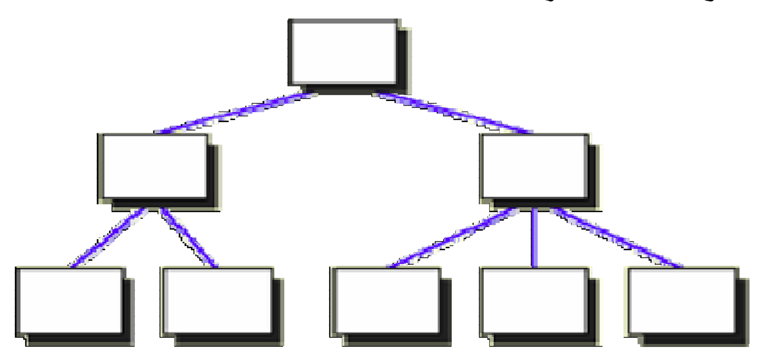


الشكل (١ –٣) نموذج الويب لتنظيم المعلومات

ويعتمد مستخدمو الويب في ذلك على أسلوب التصفح واستخدام الروابط الفائقة في استرجاع المعلومات داخل هذا النموذج. غير أن ذلك يتطلب من مستخدمي الويب مهارات تذكر الصفحات وكفاءة التحرك بينها إلى جانب الانتقال اليسير بين مستويات عرض المعلومات داخل صفحات الموقع الواحد للتعرف إلى مختلف المعلومات المتاحة. وأخيرًا فإن بنية الروابط الفائقة وعدم التزامها بمسارات شجرية أو متوازية يجعل من الصعب صيانتها وتحديثها. ويتمثل هذا النموذج في شكل صفحات الويب ومواقعها على اختلاف تخصصاتها أو وسائط حمل المعلومات بها.

٢/١/٢/١ النموذج الشجري أو الهرمي The Tree Model:

يتمثل النموذج الشجري لتنظيم المعلومات في بناء دليل يضع المعلومات المتعلقة بالموضوع أو موقع الويب في مستويات ترتب من العام إلى الخاص فالأكثر تخصصًا. وهو ما يسهل على مستخدمي الويب تحديد موضعهم داخل موقع الويب أثناء بحث واستدعاء المعلومات.



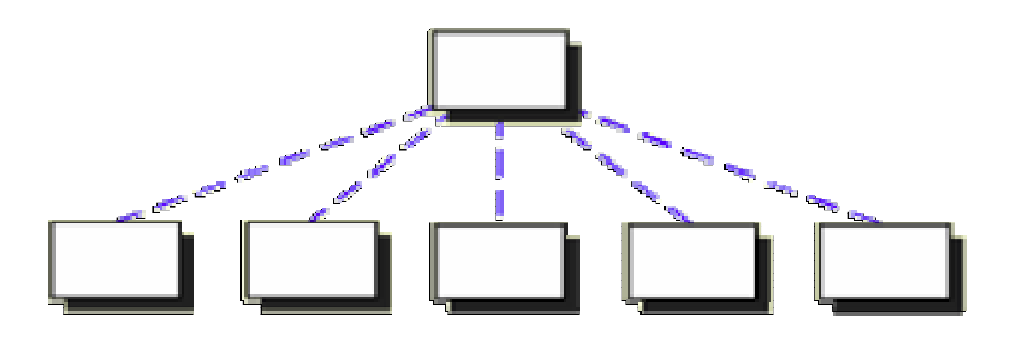
الشكل (١ –٤) النموذج الشجري لتنظيم المعلومات

تعتمد مواقع الويب في هذا النموذج على طبيعة الموضوع ذاته وتفريعاته في تنظيم معلومات الويب؛ حيث تتميز بعض قطاعات المعرفة بوجود قوائم لرؤوس الموضوعات المعيارية المحددة لطبيعة المجال أو الموضوع تحصر مختلف موضوعاته الفرعية. غير أن هذا النموذج يحتاج نوع من مستخدمي الويب يتميز بالتخصص الموضوعي إلى جانب قدرته على التحرك بين تفريعات الموضوع العامة والخاصة. ويمثل هذا النموذج أدلة بحث الويب search directories التي تقسم المعرفة أو قطاعات منها إلى فروع أخص يمكن لمستخدمي الويب التحرك بين فروع المعرفة أو القطاع المعرفي الواحد.

٣/١/٢/١ نموذج بحث المعلومات The Searchable Model:

أفرز التقدم العلمي في أجهزة وبرامج الويب ما يسمى بنموذج بحث المعلومات. وفي هذا النموذج لا يحتاج مستخدم الويب التنقل بين قطاعات الموضوع أو فرعه،

وإنما الاعتماد على استدعاء المعلومات المطلوبة مباشرة . ويعد مصطلح محركات البحث هو المسمى الآخر لنموذج تنظيم معلومات الويب؛ حيث يحوي محرك البحث قاعدة تضم نسخًا وتسجيلات من صفحات ومواقع الويب يتم استرجاعها باستخدام الكلمات المفتاحية. ومن أمثلة ذلك **Google** و**AltaVista**.



الشكل (١-٥) نموذج بحث المعلومات لتنظيم الويب

ويتصف هذا النموذج بمجموعة من السلبيات التي تعوق تحقيق الفائدة الكاملة منه، وهي:

١- يهتم هذا النموذج غالبا بتحقيق نسبة الاستدعاء أكثر من الاهتمام بنسب التحقيق للنتائج المسترجعة. مما سبب ضياع وقت وجهد مستخدمي الويب في استرجاع المعلومات.

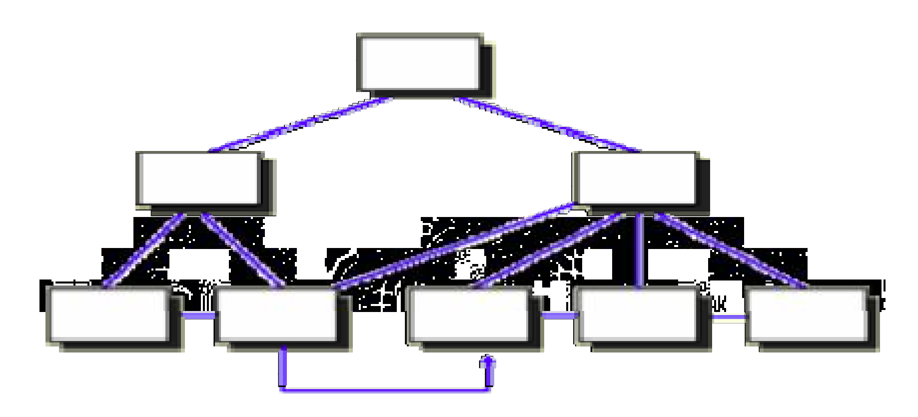
٢- يستطيع هذا النموذج إجادة التعامل فقط مع ملفات المعلومات المصاغة بلغة html دون غيرها من أنواع الملفات الأخرى، مما يتسبب معه ضياع كم هائل من المعلومات في نطاق اهتمام مستخدمي النموذج.

٣- يعتمد هذا النموذج في ترجيح المعلومات المسترجعة على حساب تكرار وتردد بعض الكلمات الواردة في صفحات ومواقع الويب، غير أن ذلك يتسبب في أحيان كثيرة في استرجاع صفحات ومواقع لا ترتبط بموضوع البحث.

٤- يحتاج نموذج محركات البحث إلى ضبط الدلالة بين كلمات العناوين في صفحات وموقع الويب وبين ما تدل عليه هذه الكلمات من موضوعات.

٤/١/٢/١ نماذج المعلومات المتداخلة أو المهجنة The Hybrid Models:

تحاول بعض مواقع الويب الجمع بين أشكال النماذج السابقة في نظام متداخل أو مهجن؛ حيث تستخدم تقسيمًا موضوعيًا يساعد مستخدمي الويب على التعرف إلى جوانب الموضوع العامة والخاصة، أيضًا استخدام الوصول السريع ببحث الكلمات الدالة، بالإضافة إلى إمكانية تصفح الموقع باستخدام الروابط الفائقة.



الشكل (١ -٦) نموذج بحث المعلومات المتداخل أو المهجن

إن استخدام هذا النموذج يتطلب من مصممي مواقع الويب دقة صياغة وتكامل أساليب تنظيم وبحث المعلومات فيما بينها، إلى جانب توفير نقاط المساعدة لمستخدمي الويب في الانتقال من أسلوب إلى آخر. وأفضل الأمثلة على هذا النموذج مواقع بوابات الويب (gateways) web portals.

٢/٢/١ المعالجة والتنظيم لمصادر معلومات الويب:

تحتاج مختلف مصادر المعلومات الرقمية على الويب إلى عمليات المعالجة والتنظيم حتى يتم التمكن من التعامل معها تنظيمًا واسترجاعًا، ويتم تناول المعالجة الفنية لمصادر الويب كما يلي:

١/٢/٢/١ معالجة صفحات الويب وتنظيمها:

ترتقي صفحة الويب إلى رتبة الموقع في حين تقف بمفردها كمصدر مستقل حمل بيانات الويب على اختلاف أشكالها ؛ حيث تصبح في هذه الحالة موقعًا على الويب من صفحة واحدة. وتتلقى صفحات الويب المعالجة الفنية وبنية التنظيم منذ عملية إنشاء وتصميم صفحة الويب، وسوف يقصر الحديث هنا على معالجة صفحات الويب في شكلها الطبيعي الذي تكون فيه وحدة متكاملة مع صفحات أخرى لتكوين موقع الويب.

هناك طريقتان يمكن بهما معالجة صفحات الويب وتنظيمها ، وهما :

الطريقة الأولى : المعالجة الفنية:

تستخدم الويب عناصر الميتاداتا metadata لوصف ملفاتها ومصادرها الرقمية عامة ومعالجة صفحات ومواقع الويب خاصة. وقد لاحظ الباحث مما شاهده خلال استخدام الويب أن المعالجة الفنية لملفات الويب تأتي بتخصص أدوات التنظيم والبحث ؛ حيث تهتم أدوات بحث الويب من محركات وأدلة بواحد أو أكثر من ملفات الويب محددة الشكل، فتوجد محركات بحث الويب ومحركات الوسائط المتعددة على اختلاف أشكالها. وعلى ذلك فإن المعالجة الفنية لصفحات الويب تدخل ضمن نطاق أدوات بحث الويب. إلا أن صفحات ومواقع الويب لم تلق الاهتمام الكافي بصياغة معايير المعالجة الفنية داخل أدوات البحث كما حظيت بها الملفات الرقمية على اختلاف وسائط حمل المعلومات.

تبدأ المعالجة الفنية لصفحات الويب مع بداية كتابة وتصميم تيجان بناء صفحة الويب html tags ؛ حيث تصاغ بعض حقول وصف الصفحة ومحتواها بعناصر مثل meta key word , meta title بما يعكس معه محتوى صفحة الويب. غير أن هذه العناصر دائمًا ما تحمل أهداف واحتياجات مصمميها لتحقيق أكبر

عدد من زيارات مستخدمي الويب للصفحة، بما يدفعهم إلى وضع كلمات ذات دلالة على موضوعات أخرى أكثر جاذبية لمستخدم الويب وإن بعدت أو حتى اختلفت تماما عن موضوع صفحة الويب الأساسي. يأتي بعد ذلك دور محركات وأدلة البحث في معالجة صفحة الويب ؛ حيث تعتمد محركات البحث بشكل أكبر من الأدلة على استخدام تيجان html tags في تنظيم صفحات الويب داخل قاعدة بياناتها. أيضًا يؤدي متن صفحة الويب دورًا مهمًا في اقتباس عناصر معالجة صفحة الويب داخل محركات البحث ، ذلك لاعتماد محركات البحث على آلية التنظيم والمعالجة بموقع الكلمات ومدى ترددها أو تواترها بكثرة داخل صفحة الويب. على الجانب الآخر ، فإن أدلة البحث تنتهج استخدام العنصر البشري في رسم الصورة الحقيقية للتعبير عن المحتوى الموضوعي لصفحة الويب ؛ حيث تصاغ الكلمات الدالة المعبرة عن الموضوع الحقيقي لصفحة الويب، إلى جانب ربط صفحة الويب بقائمة رؤوس الموضوعات داخل دليل البحث ليتم استدعاؤها عن طريق تصفح تفريعات الموضوع داخل دليل البحث.

وقد أكدت بعض الدراسات مثل Edward T O'Neill أن استخدام عناصر الميتاداتا في ازدياد مستمر داخل صفحات ومواقع الويب. وعملت هذه الدراسة على قياس نسبة استخدام حقول الميتاداتا داخل مواقع وصفحات الويب في الفترة من ١٩٩٨– ٢٠٠٢م وقد عكست نتائج الدراسة أن استخدام حقول المعالجة الفنية دائمًا في تزايد مستمر وأوضحت النتائج سعي صفحات ومواقع الويب إلى الاستعانة بمختلف معايير الوصف على الويب مثل Dublin core لوصف ملفات الويب. وذلك كما هو موضح في الجدول (١ – ٢) [١] :

(١) O'Neill, Edward T. Trends in the evolution of the public web, D-Lib Magazine, April 2003, cited 20/6/2007, cited at http://www.dlib.org/dlib/april03/lavoie/04lavoie.html

الجدول (١ -٢) نسب استخدام الميتاداتا داخل صفحات ومواقع الويب

Table 1: Metadata Usage on Public Web Sites, 1998 - 2002					
	1998	1999	2000	2001	2002
No. of Public Sites	1,457,000	2,229,000	2,942,000	3,119,000	3,080,000
% of Sites with Metadata on Home Page	70 percent	79 percent	82 percent	85 percent	87 percent
Mean No. of Tags	2.75	2.32	2.72	2.97	3.14
% of pages with Metadata	45 percent	50 percent	59 percent	62 percent	70 percent
Mean No. of Tags	2.27	2.16	2.46	2.63	2.75

الطريقة الثانية : تنظيم صفحات الويب:

إن الطريقة المثلى في تنظيم صفحات الويب هي الربط الفائق hyper linking، خاصة عندما تكون صفحة الويب جزءاً من موقع يضم مجموعة كبيرة من صفحات الويب. ومن ثم يكون التنظيم العشوائي واستخدام النص الفائق هو أحد أسلوبين لتنظيم صفحات الويب. يأتي الأسلوب الثاني متمثلاً في عمل تسجيلات قواعد بيانات محركات البحث أو أدلة البحث على الويب، حيث تنظم صفحات الويب بحسب تيجان html الواصفة للمحتوى والصفحة. وتكون هذه التسجيلات أكثر إحكامًا وتنظيمًا في أدلة البحث عنها في محركات البحث؛ حيث الوصف البشري المعتمد على معيار ثابت لوصف الويب حتى وإن كان معيارًا محلي الاستخدام in home use داخل دليل البحث.

٢/٢/٢/١ معالجة مواقع الويب وتنظيمها:

تحتاج مواقع الويب إلى عمليتي المعالجة الفنية المتمثلة في حقول الوصف الفني المعروفة بالميتاداتا، كما تحتاج أيضًا إلى جوانب التنظيم والاسترجاع المعتمدة على التقسيم الموضوعي أو البحث الحر بالكلمات المفتاحية. وتتم معالجة مواقع الويب كما يلي :

أولاً: المعالجة الفنية:

تعاني مواقع الويب من افتقارها لمعايير المعالجة والتنظيم، سواء أكان ذلك في بداية التصميم أم في داخل محركات البحث. حيث ترصد شركة Website-Designer.net أنه من بين مليار موقع على الويب تمكنت أدوات بحث الويب من الوصول إلى ٥٪ فقط، إلى جانب ٥٪ تم تكشيفها باستخدام معايير معالجة فقيرة بعض الشيء. وقد رأت الشركة أيضًا بعض جوانب الاتفاق في عناصر المعالجة بين المطبوعات ومواقع الويب، مثل اختيار المادة المكشفة، الوصف الشكلي للمصدر، طبيعة التكشيف في المنفردات أو السلاسل، طبيعة المداخل الموضوعية(١).

يرى الباحث أن هناك تداخلاً في واقع التطبيق بين مصطلحات التكشيف والوصف الفني والمعالجة والتنظيم في البيئة الرقمية عنها في البيئة الورقية التقليدية ؛ حيث يختص كل مصطلح من هذه المصطلحات بأداء عملية مختلفة تمامًا عن الأخرى في البيئة التقليدية، غير أن هذا ليس هو الوضع بالنسبة للوصف المادي والموضوعي للمصادر الرقمية؛ حيث تخضع كامل حقول التسجيلة الرقمية للبحث والاستدعاء بواسطة الكلمات المفتاحية. مثال على ذلك التكامل في استراتيجية البحث بين كلمتي " الحجم ١٦ k.b والمياه" عند استرجاع ملف لصورة رقمية أحد حقول الوصف المادي له هو الوزن ١٦ k.b ورأس موضوعه يتعلق بالمياه. وعلى ذلك فإن المعالجة الفنية والتنظيم لمواقع الويب وصفحاتها تعتمد كلية على استخدام مجموعة من الكلمات الواصفة مكونةً حقول التسجيلة الرقمية التي تخضع لآلية بحث الكلمات المفتاحية. ويعكس حقيقة ذلك اعتماد مواقع الويب ومصمميها على محركات البحث بشكل تام في وضع المعالجة الفنية لمواقع الويب داخل قواعد بياناتها ، ومن ثم تعد محركات البحث هي أدوات المعالجة والتنظيم لمواقع وصفحات الويب.

لقد أجرت جامعة University of North Carolina دراسة على عينة عشوائية لمواقع الإنترنت بلغت ٢٧٠٠ موقعٍ تم اختيار ٤٦٤ موقعًا لدراسة ملامح التسجيلات

(١) Website Indexing. Website-Designer, 2007, cited at 6/6/2007, cited at http://www.website-designer.net/main.php

الفنية بها، وقد أكدت الدراسة على افتقار مواقع الويب عامة لمعايير المعالجة الفنية اللازمة لعمليات التنظيم والاسترجاع. كما عملت الدراسة على تحديد الأطر العامة التي تتسم بها تسجيلات الوصف الحالية، وخرجت نتائج الدراسة بأحد عشر ملمحًا أدرجت أسفل ثلاثة قطاعات لوصف مواقع الويب وصفاً فنيًا. يوضحها الجدول (١ -٣). بالإضافة لذلك نبهت الدراسة القائمين على معالجة المواقع بأهمية الواقع التفاعلي والديناميكي لمواقع الويب، حيث يتسم محتوى مواقع الويب بالتغيير والتحديث الدائم، مما ينعكس على تنظيم حقول التسجيلات الرقمية[١]:

الجدول (١ -٣) ملامح وجوانب الوصف لمواقع الويب

Facet	Category
Content	Website Navigation: issues involved in navigating and accessing the website
	Categorization of Geographic Scope : issues involved in defining the geographic scope of the website
	Categorization of Topical Scope: issues involved in defining the topical scope of the website
	Miscellaneous: about issues related to website cataloging that fall into none of the above categories
Format	Question: Any comments that are questions along with anything that can reasonably be inferred to be a question.
	Answer: Any comment explicitly in response to a question and comments which probably answer unasked questions
	Statement: Declarative statements
Function	Log of Action: A statement of an action taken in the past
	Reminder: A statement to remind catalogers of actions that should or should not be taken in the future and relevant information that they should notice in the future

(١) Luo, Lili and West, David and Marchionini, Gary. A study of annotations for a consumer health portal , 2006, cited at 20/6/2007, cited at http://ils.unc.edu/annotation/publication/luo_jcdl_2005.pdf

Facet	Category
	Reach Consensus: A statement made in the process of reaching an agreement on a disputed point.
	Action Request: A comment that request a cataloger to take an action or provide information

يوضح الجدول (١ – ٣) القطاعات الأساسية الثلاث لبناء التسجيلات الرقمية لمواقع الويب، وقد شملت كل منها العناصر المتعلقة بكل جانب من جوانب الوصف. وتعكس هذه العناصر الملامح التالية:

أولاً: بنيت التسجيلة الرقمية على عناصر ثلاث هي المحتوى content والشكل format والمهام أو الوظائف Function. وقد عكس ترتيب هذه العناصر في الجدول الاهتمام الكبير بمحتوى موقع الويب، والرسالة المعلوماتية المطروحة داخل المحتوى الذي يعتبر وسيلة التنظيم والاسترجاع الأساسية لمواقع الويب. أعقب ذلك الحديث عن الشكل ثم المهام أو الوظائف الفنية والتقنية لموقع الويب.

ثانيًا: كان أول عناصر التسجيلة هو المحتوى المعبر عن الاتجاهات الموضوعية المختلفة لمواقع الويب، وظهر ذلك في عناصر :

١- Website Navigation التصفح : ويتضمن التقسيم الموضوعي الذي بني عليه موقع الويب.

٢- Categorization of Geographic Scope تصنيف المجال الجغرافي : الجانب المكاني أو الجغرافي الذي يركز عليه محتوى موقع الويب.

٣- Categorization of Topical Scope تصنيف المجال الموضوعي : ويغطي الموضوع أو القطاع المعرفي الآتي في متن أو وسائط موقع الويب.

٤- Miscellaneous المتنوعات : يغطي هذا الحقل الجوانب الموضوعية التي لم تذكر في العناصر السابقة ويرى واصف موقع الويب أهميتها في التعبير عن محتوى الموقع.

ثالثًا: لم يحتوِ عنصر الشكل format على خصائص الوصف المادي كما نراه في تسجيلات الوصف الفني عامة، غير أن ذلك يتفق واقعيًا مع موقع الويب كمصدر معلومات لا توجد له خصائص أو ملامح مادية. واقتصر عنصر الشكل هنا على ذكر أهم التعليقات والحقائق عن الموقع التي يمكن أن تطرح كأسئلة بحث تستخدم في استرجاع موقع الويب.

رابعًا: اهتم العنصر الثالث المهام Function بتاريخ ومستقبل موقع الويب من الأحداث أو التغييرات التي مر أو سيمر بها موقع الويب، ويعكس ذلك عنصر Log of Action المخصص لسرد تفاصيل التغييرات التي مر بها الموقع، وعنصر Reminder لتحديد التغييرات التي لابد من إحداثها للموقع في المستقبل، وعنصر Reach Consensus الذي يقترب من الحديث عن المسؤولية والملكية الفكرية لموقع الويب، وختاما كان عنصر Request Action الذي يشمل القائم على معالجة ووصف موقع الويب؛ حيث يذكر فيه أهم التعليقات والخصائص التي يراها في الموقع.

ثانيًا: تنظيم مواقع الويب:

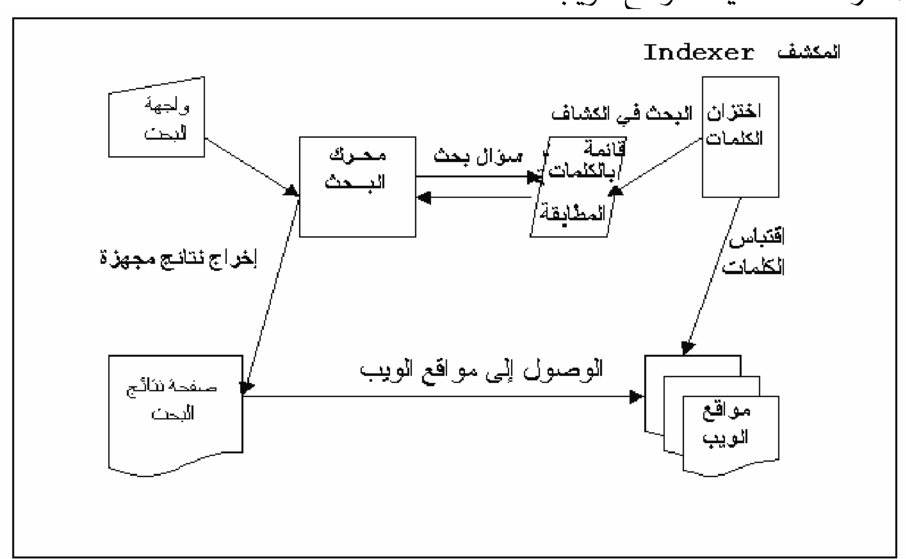
تعد محركات وأدلة بحث الويب وحدها الأداة المنوطة بتنظيم مواقع الويب ومختلف الملفات الرقمية الأخرى. وعندما نتناول تنظيم مواقع الويب فإننا نتحدث بالفعل عن عملية التكشيف والتحليل داخل محركات وأدلة البحث؛ حيث تستخدم محركات البحث أحد برامجها " المكشف Indexer" لتحليل محتوى ونص مواقع وصفحات الويب باستخدام آليات مختلفة لاستخراج الكلمات الدالة المعبرة عن محتوى موقع الويب ؛ فتعتمد محركات البحث مثالاً على عدد تردد كلمات محددة في نص صفحة الويب أو ظهور بعض الكلمات في أماكن التعريف بالمحتوى داخل صفحة الويب مثل تيجان meta key word, meta content. وعلى ذلك تقف الدراسات عند تناول عملية التنظيم والتحليل داخل محركات

وأدلـة البحـث دون الحـديث عـن تنظيم مواقـع الويب ذاتهـا ، ذلـك أن بدايـة عمليـة التنظيم ونهايتها تتم داخل قاعدة بيانات أدوات البحث باستخدام البرامج الآليـة أو العنصر البشري.

تمثل عملية تنظيم وتحليل محتوى مواقع الويب واسطة العقد بين عمليات ثلاثة تتم داخل أدوات بحث الويب. وتتم عملية التنظيم لمحتوى مواقـع الويب مـن خـلال تكشيف مـتن صـفحات الموقـع والخـروج بـأهم الكلمـات الدالـة علـى موضـوع أو موضوعات موقع الويب. فبعـد أن تنتهي مرحلـة إضـافة نسـخة مـن موقـع الويـب إلى قاعدة بيانـات محـرك البحـث ، يبـدأ عمـل البرنـامج الثـاني في أدوات البحـث الـذي يعرف بالمكشف أو المفهرس ، حيث يقوم بإعداد تسجيلات ببليوجرافية لمواقـع الويب حديثة الإضافة إلى قاعدة بيانـات أداة البحـث ، وتعمـل هـذه التسجيلة وفقًـا لما ورد في المكتبات والمعلومات عن (التمثيل المكثف) حيث لا يتم اختزان موقـع الويب بأكمله في عدد مـن أدوات البحـث ، وإنمـا يختـزن الموقـع في شـكل حقـول وبيانات مثل ؛ مسـار هـذا الموقـع للـربط بـه والوصـول إليـه ، وكـل ذلـك في شـكل تسجيله كما هو الحال في الفهارس الآلية للمكتبات OPAC.

وتـتم عمليـة الفهرسـة في حالـة الأدلـة بشـكل يـدوي ، أمـا في حالـة محركـات البحث فإنها تتم آليًا ، وفي حالة التكشيف والفهرسة الآلية فإن برامج التكشيف الآلي تستخدم بنية صفحة الويـب في التعـرف إليهـا ووضـع كلمـات أو مـصطلحات معـبرة عـن الموضـوع أو الموضـوعات الـتي تـضمها صـفحة الويـب؛ حيـث يفحـص البرنـامج كـل أجـزاء الـصفحة الـتي تحتـوي علـى كلمـات تخـص الموضـوع مـن كلمـات العنـوان ، وكلمـات الـنص ، والكلمـات المفتاحيـة الـتي وضـعها مـصمم صـفحة الويـب في رمـوز meta tags وتجمـع هـذه الرمـوز كـل الكلمـات الـتي يـرى مصمم الصفحة أنها تعبر عن موضـوع أو موضـوعات صـفحة الويـب. وبعـد تفحـص كـل الجوانـب الـسابقة يتم وضـع كلمـات مفتاحيـة داخـل تـسجيلة هـذا الموقـع أو وضـع كـل كلمـات الـنص الطبيعيـة في وضـع نـشط ليـتم مضاهاتها بعـد ذلـك مـع

كلمات البحث التي يضعها المستفيد عند إجراء البحث. وتختلف مواضع التكشيف من برنامج إلى آخر ؛ حيث تعمل بعض برامج التكشيف على الاكتفاء بالجزء الأول فقط من بنية اللغة المعيارية لصفحة الويب، كما تعمل بعض البرامج على تكشيف عناوين الصفحات إلى جانب تكشيف مسار الصفحة URL أيضًا، وتمتد قدرة برامج التكشيف في أحيان أخرى إلى تكشيف كامل متن صفحة الويب وهو ما يجعل محركات البحث في هذه الحالة تمتاز بوفرة في الاستدعاء مع انخفاض نسبة التحقيق. ويبين الشكل (١-٧) مرحلة تكشيف مواقع الويب[١]:



الشكل رقم (١-٧) مرحلة تكشيف مواقع الويب

وقد أشارت Laura Cohen إلى بعض الجوانب التي تؤخذ في الاعتبار عند تنظيم مواقع الويب، مثل[٢]:

---

(١) سيد ربيع سيد. محركات بحث الصور على الإنترنت : دراسة تحليلية لوضع مواصفات محرك بحث صور... ، إشراف هانئ محيي الدين عطية ، جامعة بني سويف، ٢٠٠٥م، أطروحة ماجستير، ص ٢٠.

(٢) Cohen , Laura. Understanding the world wide web. Internet tutorials, 2007, cited 20/6/2007, cited at http://www.internettutorials.net/

❖ تغير بعض مواقع الويب بسرعة في المحتوى نظرا للتحديث السريع، ولذلك تتغير الكلمات الدالة على أثر تغيير المحتوى.

❖ تعطي إمكانات النشر، التي توفرها حزمة برامج المكتب office، مواقع الويب صلاحية النشر غير المقنن بمختلف الأشكال، مما يؤثر بدوره على تقنين عملية معالجة وتنظيم مواقع الويب.

❖ إن الوصول إلى سياسات النشر لدى مواقع الويب غالبًا ما يعتمد على المصادفة ؛ حيث تحرص مواقع الويب على الاحتفاظ بسياسة نشر وتصميم مواقع الويب حتى تحتفظ بتميزها أو خصائصها عن المواقع الأخرى.

وقد أضافت الكاتبة أيضًا بعض المراحل المقترحة لتنظيم مواقع الويب، يمكن استخدامها كخطوات تقف عليها معالجة مواقع الويب:

١- تحديد أي مواد المعلومات هي الأهم في مواقع الويب للتعبير عن محتواها أو موضوعاتها.

٢- الوقوف على الكلمات الدالة في التعبير عن تلك المواد.

٣- اتباع ملامح استراتيجية التنظيم التي يعتمد عليها في تنظيم مواقع الويب الحالية والمصادر المستقبلية.

تنتهي عملية تنظيم محتوى مواقع الويب عند قائمة بالمصطلحات الموضوعية مرتبة تنازليًا من العام إلى الخاص فالأكثر تخصصًا داخل أدلة بحث الويب، وقد ساعد في إعدادها بهذا الشكل مجموعة من العناصر البشرية القائمة على عملية تحليل محتوى مواقع الويب. أيضًا توجد النتيجة نفسها مع الاختلاف في الشكل داخل محركات بحث الويب ؛ حيث تنظم مواقع الويب داخل قواعد بيانات المحركات بالاعتماد على الكلمات الدالة وخصائص الكشاف المقلوب inverted index ، وقد جاءت هذه النهاية بأداء برنامج التكشيف داخل محرك البحث.

ثالثًا: تنظيم المحتوى التفاعلي الديناميكي لمواقع الويب:

مما سبق يتضح أن مواقع الويب تتميز بسرعة تغيير المحتوى والتحديث الدائم. وكان لذلك أثره على تنظيم محتوى مواقع الويب ؛ حيث خرجت بعض الأفكار التي تدعو إلى استخدام خرائط مواقع الويب web sites maps ، ويتجلى دور هذه الخرائط في رسم الواقع الحقيقي والعملي لموقع الويب من خلال متابعة شبكة المسارات داخل بنية الموقع Site URLs structure. يعكس ذلك طبيعة ترابط صفحات الويب واتجاهاتها ، أيضًا متابعة نشاط أو انكسار بعض مسارات الموقع في عمليات التحديث المختلفة. وباختلاف فترات رسم هذه الخرائط لمواقع الويب يختلف التعبير عن شكل المحتوى ومن ثم اختلاف المصطلحات الكشفية الدالة على المحتوى. وتوفر البيانات الناتجة من وصف المحتوى مجموعة من التسجيلات records المترابطة عن محتوى كل صفحة داخل موقع الويب. ويعد الترابط بين التسجيلات باستخدام الروابط الفائقة هو أهم الصفات اللازم توافرها في المعالجة الفنية لمواقع الويب. ويمكن أن تتسم هذه التسجيلات ببعض الضوابط، مثل:

- إظهار هذه التسجيلات للمشتركين أو المسجلين داخل قائمة موقع الويب.
- إخفاء هذه التسجيلات بعيدًا عن أعين مستخدمي الويب مع الاحتفاظ بإمكانية إظهارها عند الحاجة.
- الإخفاء التام لهذه التسجيلات أمام مستخدمي الويب وإظهارها فقط أمام مسئولي الموقع واحدة من خصائص المسؤولية عن موقع الويب.

٣/٢/٢/١ معالجة وتنظيم قواعد بيانات الويب:

تختلف قواعد بيانات الويب عن غيرها من مصادر الويب في المعالجة الفنية والتنظيم. حيث تعالج صفحات ومواقع الويب المعالجة الخارجية التي تشمل الحديث عن وصف وتنظيم الموقع كاملاً كوحدة واحدة. غير أن المعالجة الفنية لقواعد البيانات تهتم بالمعالجة الداخلية لملفاتها واتباع طرق التنظيم الملائمة

لطبيعة ملفات كل قاعدة بيانات على حدة. وبشكل آخر فإن المعالجة الفنية لمواقع وصفحات الويب تهدف إلى تنظيم المواقع معا، بينما تهدف المعالجة الفنية لقواعد البيانات إلى تنظيم مجموعة ملفات كل قاعدة على حدة وليس مجموعة قواعد بيانات الويب معا. غير أن هذا الواقع تغير الآن بالحديث عن تنظيم قواعد بيانات الويب معًا من خلال أدوات تنظيم وبحث تتعامل مع قواعد البيانات على مستوى قاعدة البيانات كوحدة وليس على مستوى الملف كما ألِفَ مستخدمو الويب منذ نشأة الويب. وأصبح الاتجاه الآن نحو تقنين المعايير الملائمة لوصف قاعدة البيانات كمصدر للمعلومات على الويب، وهو ما يعد من المشكلات الفنية الآنية التي تشغل مصممي أدوات بحث الويب. ولهذا جاءت دراستنا الحالية.

٣/١ أنماط نظم استرجاع الويب:

تحوي الويب مجموعة متنوعة من أشكال نظم استرجاع المعلومات التي تختلف في خصائصها ومستوياتها، مثل:

١/٣/١ نظم استرجاع صفحات الويب:

حددت Laura Cohen أساليب استرجاع صفحات الويب في أربعة أشكال هي[١]:

١- استخدام مسار صفحة الويب URL داخل فراغ العنوان متصفح الويب.

٢- استخدام أسلوب التصفح اعتمادا على الروابط بين صفحات الويب.

٣- البحث داخل إحدى القوائم الموضوعية لأحد أدلة بحث الويب المرتبطة بمجموعات الصفحات.

٤- استخدام أحد محركات بحث الويب بإدخال كلمات مفتاحية لاسترجاع المطابق لها من صفحات الويب.

ويمكن تناول مختلف نظم استرجاع صفحات الويب السابقة كما يلي:

(١) Cohen , Laura. Ibid.

١/١/٣/١ نظام الاسترجاع باستخدام URL:

تعتمد الويب في بنائها على استخدام الروابط الفائقة للربط بين صفحات الموقع الواحد والمواقع المختلفة. ولذلك فان أي ملف أو صفحة أو مصدر معلومات على الويب لابد أن يحمل مسارًا أو عنوانًا ثابتًا لا يتطابق مع أي مصدر آخر على الإطلاق. والتسمية الشائعة لهذا العنوان هو URL أو محدد المصادر الموحد. وتخصص برامج تصفح الويب على اختلافها مكان للبحث بإدخال مسار المصدر يسمى Address Bar أو شريط العنوان. ولكي تتعرف برامج التصفح على هذا المصدر واسترجاعه أمام مستخدمي الويب، فإن هناك صيغة موحدة ثابتة لكتابة مسار ملف أو مصدر الويب تعتمد على أركان ثابتة، يوضحها الجدول (١ – ٤) ومن ثم يستطيع أي مستخدم استدعاء موقع الويب بشكل محدد عند كتابة المسار بشكل صحيح. غير أن هذا الأسلوب من الاسترجاع يتطلب من مستخدمي الويب حفظ مسارات مواقع الويب موضع الاهتمام وهو ما لا يمكن لأي مستخدم القيام به.

الجدول (١–٤) مكونات مسار مواقع الويب

الوظيفة	عناصر المسار
بروتوكول النص الفائق	**Protocol:** http
نوع خدمة الويب والخادم المضيف للموقع	**Host computer name:** www
اسم موقع الويب	**Second-level domain name:**
نوع وطبيعة موقع الويب	**Top-level domain name:**
امتداد اسم الدليل	**Directory name:**
اسم الملف	**File name:**

٢/١/٣/١ نظام الاسترجاع باستخدام برنامج التصفح Browser:

تشكل صفحات ومواقع الويب معًا نسيجًا عنكبوتيًا متداخلاً ومتقاطع المسارات. حيث يمكن نظريًا تصفح كل صفحات الويب ومساراتها انطلاقًا من أي صفحة ويب تشكل نقطة البداية للتنقل والتصفح داخل الويب. وليس من المبالغة القول أن هذه المسارات المتداخلة تشكل معًا النسيج العصبي لجسد الويب، الذي إذا ما فقدته الويب فقدت الترابط والقدرة على التنظيم. ولا يتطلب نظام الاسترجاع بتصفح مواقع الويب سوى استخدام أحد برامج التصفح مثل internet explorer or Netscape navigator. غير أن هذا الأسلوب يأخذ مستخدمي الويب إلى أبعد مما يحتاجون من المعلومات المحددة، ذلك عن طريق ما يحدثه التصفح من تشتيت لأفكار الباحثين بمرورهم بصفحات تقترب أو تبتعد من موضوع البحث.

وتقسم برامج تصفح الويب إلى نوعين هما:

- ☒ **متصفحات رسومية** Graphical:

وهي البرامج التي تستطيع قراءة وسائط النص والصورة والصوت والفيديو. وأشهر برامج التصفح الرسومية : Internet Explorer, Netscape, Mozilla and Opera. وتعمل هذه البرامج في بيئة التشغيل الويندوز windows ، حيث يتم تحميلها بتحميل نظام تشغيل الحاسبات الشخصية.

- ☒ **متصفحات نصية** Text:

وتتصفح هذه البرامج الويب بشكل نصي يعرض فقط أسماء الملفات مع إظهارها بشكل مختلف وسط البنية العامة لصفحة الويب. ويتم التحكم في تصفح الويب باستخدام مؤشرات لوحة المفاتيح لأعلى وأسفل. وأشهر هذه البرامج Lynx.

٣/١/٣/١ نظام الاسترجاع باستخدام أدلة بحث الويب:

تسمى أدلة بحث الويب بشجرة الموضوعات subject tree. وتعتبر أدلة الويب فهارس لمعالجة وتنظيم صفحات ومواقع الويب باستخدام العنصر البشري، ومن الممكن أن تبدأ أدلة البحث بمجموعة قليلة من رؤوس الموضوعات ثم تتزايد بتزايد أعداد الصفحات والمواقع المضافة إلى قاعدة دليل البحث. وتنظم صفحات الويب بحسب تحليل المحتوى والتعرف إلى الموضوعات التي تحويها بداخلها، ومن ثم تأخذ كل صفحة ويب المصطلح الدال على هذا الموضوع مع وضع رابط بينها وبين المصطلح الدال في قائمة رؤوس الموضوعات داخل دليل البحث. وما على مستخدم الويب سوى تصفح مصطلحات قوائم رؤوس الموضوعات داخل دليل البحث والضغط على رأس الموضوع مجال البحث لاستدعاء مختلف صفحات ومواقع الويب المرتبطة به. ويمتاز دليل البحث بدقة المواقع المستدعاة لموضوع محدد، وذلك يأتي من استخدام العنصر البشري في عملية المعالجة والتحليل الموضوعي لصفحات الويب. أيضًا فإن أدلة البحث تساعد مستخدمي الويب في الوصول إلى نتائج جيدة تتعلق بالموضوعات التي ليس لديهم فكرة كبيرة عنها، خلاف محركات البحث التي تتجه إلى التخصيص نتيجة استخدام الكلمات المفتاحية في عملية البحث. ومن أشهر أدلة البحث على الويب دليل ياهو Yahoo الذي عمل مؤخرًا على إضافة آلية البحث بالكلمات المفتاحية كخيار إضافي لبحث قاعدة بياناته[١].

٤/١/٣/١ نظام الاسترجاع باستخدام محركات بحث الويب:

يمكن وضع تعريف إجرائي لمحركات البحث هو " إحدى أدوات بحث استرجاع المعلومات على الويب، وتعتمد على ثلاثة برامج أساسية في ذلك هي الزاحف والمكشف وآليات البحث، بما يوفر القدرة على الإضافة والبحث ".

(١) SUBJECT DIRECTORIES. debbie flanagan, 2004, cited at 10/6/2007, cited at http://www.learnwebskills.com/search/subject.html

وتمثل محركات البحث العمود الفقري لتنظيم واسترجاع صفحات ومواقع الويب. حيث تعمل محركات البحث نظريًا على جمع وتنظيم وبحث كل صفحات ومواقع الويب التي تضاف كل يوم إلى بيئة الويب. وتختلف المحركات عن الأدلة كنظم استرجاع في أكثر من جانب، مثل[1]:

أولاً: تختار أدلة البحث ما يضاف إليها بانتقاء العنصر البشري لطبيعة وتخصص دليل البحث، بينما يضم برنامج الزاحف في محرك البحث كل ما يواجهه من صفحات الويب بغض النظر عن طبيعتها أو محتواها.

ثانيًا: تستخدم أدلة البحث العنصر البشري في أداء عمليات المعالجة والتنظيم المختلفة، بينما يعتمد محرك البحث على برامج ثلاثة للقيام بالإضافة والتنظيم والبحث هي على الترتيب: spiders or robots , indexer and search operators.

ثالثًا: تستخدم أدلة البحث حقول الوصف والتسجيلات الرقمية لمعالجة الصفحات والمواقع، بالإضافة إلى تنظيم الصفحات والمواقع وفقًا لقوائم رؤوس الموضوعات المعيارية. وتعمل محركات البحث باستخدام المعالجة والتنظيم وفقًا للكلمات المفتاحية.

رابعًا: تستخدم أدلة البحث تصفح قوائم رؤوس الموضوعات وترتيبها من العام إلى الخاص لاسترجاع صفحات ومواقع الويب، أما محركات البحث فتستخدم الكلمات المفتاحية كأسلوب للاسترجاع.

ومن أشهر محركات البحث google, altavista and msn. وتجدر الإشارة هنا إلى أن تطور أدوات بحث الويب قد أدى إلى ظهور **البوابات** web gate way or portals. والبوابة هي أداة لبحث الويب تستخدم أسلوبي التصفح والكلمات المفتاحية في الاسترجاع. وتتميز البوابات عن الأدلة والمحركات بتقديم خدمات أخرى مثل البريد الإلكتروني ومجموعات الأخبار وخدمات النقاش.

---

(١) سيد ربيع سيد. مصدر سابق.

٢/٣/١ نظم استرجاع مواقع الويب:

تتفق مواقع الويب مع صفحات الويب في استخدام نظم الاسترجاع نفسها، غير أن موقع الويب يزيد من صعوبة عمل وأداء نظم الاسترجاع السابقة؛ حيث يخضع موقع الويب المكون من مجموعة من الصفحات دائمًا إلى التغيير في المحتوى وحجم الصفحات المترابطة داخله مما يزيد من صعوبة وصف محتوى الموقع، سواء أكان ذلك على فترات زمنية يتخللها تحديث للموقع أم كان على مستوى كل الصفحات الداخلية المكونة لموقع الويب. وتعجز بعض برامج الزاحف داخل محركات البحث عن الوصول إلى مستويات متعمقة لتحليل موقع الويب المكون من عشرات الصفحات، نظرًا لمستوى التحليل المحدود في برنامج الزاحف.

يمكن القول إن مواقع الويب تقف على خط فاصل بين سهولة المعالجة لمحتوى صفحات الويب في نظم الاسترجاع، وبين قواعد البيانات التي تمثل المشكلة الرئيسية في المعالجة والتنظيم داخل نظم الاسترجاع على الويب. وهذا يرجع إلى أن موقع الويب يحمل نفس طبيعة بناء html مثل صفحات الويب، غير أنه يحمل طبيعة عمق المحتوى وديناميكيته التي توجد داخل قواعد البيانات. ولن يتم هنا تكرار ما سبق ذكره عن نظم استرجاع صفحات الويب؛ لتماثلها مع مواقع الويب ثابتة المحتوى.

٣/٣/١ نظم استرجاع قواعد البيانات:

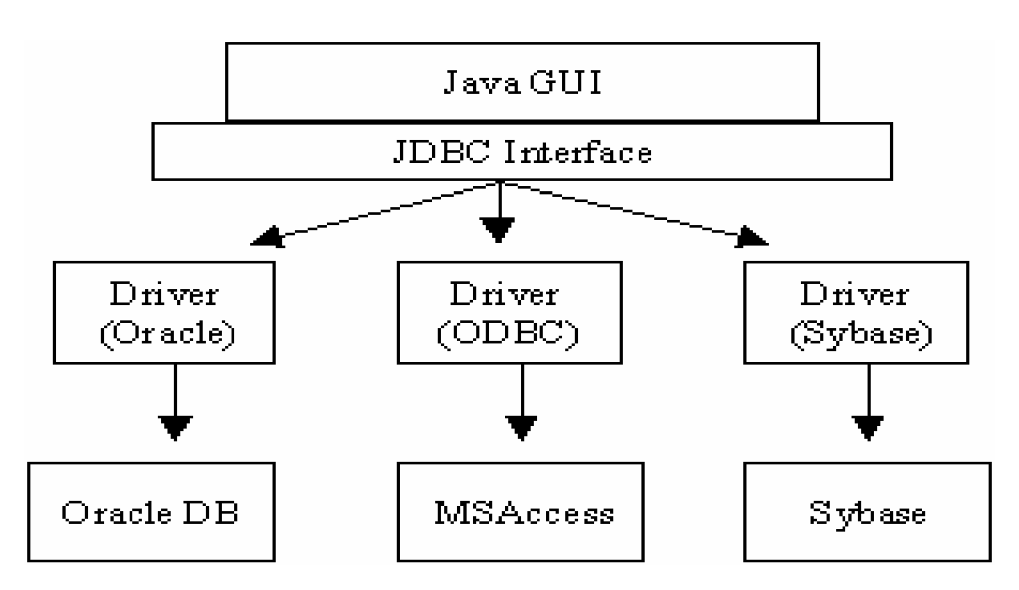
تختلف قواعد البيانات في طبيعتها عن بنية الويب؛ حيث تعتمد قواعد البيانات على نظم إدارة، مثل: MS SQL, Oracle, MS Access, My SQL. وهي بدورها لا تتعامل مع بنية النص الفائق في تنظيم ملفاتها. ويترتب على ذلك عدم استطاعة نظم استرجاع الويب الحالية التعامل مع محتوى هذه القواعد؛ لاختلاف طبيعتها عن بنية برنامجي الزاحف والمكشف في محركات البحث. ولدخول هذه المصادر الرقمية (قواعد البيانات) إلى الويب تحتاج لبنية وسيطة تجمع بين خصائص واجهات النص الفائق من جانب وآليات عمل قواعد البيانات من جانب آخر. وسميت هذه

الواجهات بـ CGI داخل خادم الويب من جانب المستخدم، بالإضافة إلى استخدام the JDBC interface من جانب قواعد البيانات. ويمثل ذلك الشكل (١-٨)[١].

إن ثمة مستويين للاسترجاع تمر بهما قواعد البيانات على الويب:

الأول : يتمثل في أساليب استرجاع الملفات الداخلية المكونة لقاعدة البيانات ويقوم بذلك النظام الداخلي لإدارة قاعدة البيانات DBMS.

والثاني : يتمثل في استرجاع قاعدة البيانات كوحدة واحدة، ثم التعامل مع ما بداخلها من ملفات وتقوم بذلك محركات وأدلة بحث الويب غير المرئية. ويمثل الجانب الثاني المشكلة التي تواجهها نظم الاسترجاع الآن على الويب، ويطلق على هذه المشكلة ظاهرة الويب غير المرئية؛ لعدم تمكن هذه النظم من رؤية محتوى قواعد البيانات والقدرة على أداء عمليتي المعالجة والتنظيم اللازمتين لاسترجاع ملفات قواعد البيانات أمام المستفيدين.



الشكل (١-٨) الواجهات الوسيطة بين الويب وقواعد البيانات

(١) Sol, Selena. Introduction to databases for the web, 1998, cited at 10/6/2007, cited at http://www.databasejournal.com/sqletc/article.php/26861_1431601_1

٤/١ أشكال محتوى الويب:

تقسم محركات البحث محتوى الويب إلى قسمين ؛ الأول : الويب المرئية وهو كل ما تستطيع محركات البحث الوصول إليه ومعالجته وتنظيمه على اختلاف أنواع ملفات المعلومات أو وسائط حملها. الثاني: الويب غير المرئية وهو كل ما لا يمكن لمحركات البحث التعامل معه أو إضافته إلى مستودعاتها. فالقدرة على الوصول للمعلومات أو عدم الوصول لها عند محركات البحث، تمثل قدرة المحركات على رؤية المعلومات الرقمية أو عدم رؤيتها. ومن ثم فإن معيار الحكم على المعلومات الرقمية بكونها مرئية أو غير مرئية داخل المحركات، هو وصول أو عدم وصول محركات البحث إليها أو عدم وصولها. ويمكن توضيح شكلي المعلومات المرئية وغير المرئية كما يلي :

١/٤/١ الويب المرئية:

يطلق على الويب المرئية مسميات أخرى مثل الويب السطحية surface web ، وهو المعنى المقابل لأحد مصطلحات الويب غير المرئية المسمى بالويب العميقة the deep web. وقد أوردت موسوعة Wikipedia encyclopedia تعريفاً للويب المرئية هو "إن الويب السطحية – التي تعرف أيضًا بالويب المرئية أو المكشفة- هي ذلك الجزء من الويب القابل للوصول والمكشف بواسطة محركات البحث. وتكون محركات البحث مستودعاتها من مصادر الويب باستخدام برامج الزاحف أو العنكبوت التي تبدأ عملها بالاعتماد على قائمة مواقع معلومة مسبقاً ، ويقوم الزاحف بالوصول لصفحات الويب ونسخها إلى مستودع محرك البحث. وتستخدم برامج الزاحف الروابط الحديثة في الوصول إلى صفحات أخرى، ومن ثم تكون كل صفحات الويب قابلة للوصول والتكشيف بواسطة محركات البحث. وفي حالة وجود عوائق تمنع برامج الزاحف من أداء عملها مثل كلمات المرور أو مستويات التحليل العميقة أو تقنيات متطورة ... فإن محرك البحث لا يرى هذه

الصفحات وتدخل بذلك ضمن ما تعرف بالويب غير المرئية[1]." ويتضمن هذا التعريف بعض الجوانب مثل:

١- الربط بين مصطلحي الويب المرئية والويب المكشفة، بما يؤكد أن الرؤية إنما هي رؤية محرك البحث لصفحات الويب حتى يتمكن من التعامل معها.

٢- اشتمل التعريف السابق على كل جوانب عمل محرك البحث حيث تكون كل صفحات الويب قابلة للوصول والتكشيف بواسطة برنامج الزاحف.

وتشمل الويب المرئية مختلف أشكال صفحات ومواقع وملفات الويب والقليل من قواعد البيانات على الويب ؛ حيث تكون هذه المعلومات متاحة أمام برامج الزاحف للتعامل معها ونسخها داخل مستودعات محركات البحث. وتستخدم برامج الزاحف تقنيات الروابط الفائقة بين صفحات الويب لتشكل نسيجًا من مصادر المعلومات يظهر بعضه بعضًا.

٢/٤/١ الويب غير المرئية:

يعرف قاموس مصطلحات المكتبات والمعلومات ODLIS الويب غير المرئية تحت مصطلح الويب العميقة على أنها " المعلومات المتاحة للوصول من خلال شبكة الويب العالمية، غير أنه لا يمكن استرجاعها بواسطة محركات البحث التي تعتمد على برامج الزاحف أو العنكبوت، مثل أشكال المعلومات التي توجد في ملفات pdf داخل قواعد البيانات التي تعتمد فقط على أسئلة البحث بشكل محدد. وتحوي الويب العميقة كمًّا من المعلومات يفوق الويب السطحية بـ٤٠٠-٥٠٠ مرة، يوجد أكثر من نصفها داخل قواعد البيانات، وتستخدم الويب بعض الخدمات لاسترجاع هذه المعلومات مثل CompletePlanet and ProFusion."[2]

ويظهر التعريف السابق بعض الجوانب مثل :

(١) Surface Web. Wikipedia : the free encyclopedia, 2007, cited at 20/6/2007, cited at http://en.wikipedia.org/wiki/ Surface_ Web

(٢) ODLIS —Online Dictionary for library and information science. OP.cit

١- إن الهدف من المعلومات الرقمية هو الوصول إليها من خلال محركات البحث التي تعد نافذة مستخدمي الويب لرؤية مصادر المعلومات الرقمية.

٢- إن السبب الرئيس لظهور الويب العميقة هو وجود المعلومات الرقمية في أشكال متنوعة مثل ملفات ...pdf , mdb, ppt.

٣- إن المصدر الرئيس لتكوين الويب غير المرئية هو قواعد البيانات حيث تتصف بخصائص مخالفة لطبيعة العمل داخل محركات البحث بما يصعب على المحركات الوصول إلى محتواها.

ويرى الباحث أن الويب المرئية تتكون من مجموعة من المصادر تعتمد في بنيتها الأساسية على بنية html التي تستخدمها محركات البحث للتعرف إلى المصادر حديثة الإضافة، إلى جانب ما بداخل هذه الصفحات من ملفات ذات روابط محددة يمكن لمحركات البحث المتخصصة الوصول إليها بحسب نوع الملفات، وذلك خلاف ما تكون عليه قواعد البيانات. ويرى الباحث أن مصطلح قواعد البيانات بمثابة المرادف الموضوعي لمصطلحات الويب غير المرئية والويب العميقة والويب الخفية. ويوضح هذه الفكرة واقع العمل بين محركات البحث وقواعد البيانات ؛ حيث تستطيع محركات البحث الوصول إلى أي مصدر معلومات دون قواعد البيانات لما لها من طبيعة بناء وتصميم لا تتوافق مع محركات البحث في العمل ومن ثم يشكل محتوى هذه القواعد بيئة جديدة غير مرئية أمام محركات البحث.

وقد أورد الباحث في دراسته للماجستير تفصيلاً لمستويات وأنواع قواعد بيانات الويب، حيث يمكن على أثرها تقسيم مستويات الوصول إلى معلومات الويب ابتداءً بالمعلومات ذات الرؤية غير الشفافة ثم انتهاءً بالشكل غير المرئي تمامًا لقواعد البيانات. وكانت مستويات رؤية المعلومات كالتالي[1]:

---

(١) سيد ربيع سيد. محركات بحث الصور الثابتة : دراسة تحليلية.- ط١.- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧م.-ص ١٢٠.

١/٢/٤/١ الويب غير الشفافة:

وهي تشمل قواعد البيانات كبيرة الحجم على الويب، التي تتكون بدورها من مجموعات كبيرة من مواد المعلومات، وعلى الرغم من ملاءمة هذه المواد لعمل برنامج الزاحف وإمكانية اقتنائها داخل قواعد محركات البحث، إلا أن برامج الزاحف لا تستطيع الوصول إليها ويرجع Paul Pedley ذلك إلى عوامل مختلفة منها:

- عمق عملية الزحف: تعمل محركات البحث على زيادة عمق التحليل والبحث لبرامج الزاحف لتحقيق أفضل أشكال الاقتناء، خاصة في الوقت الذي تتزايد فيه أعداد صفحات مواقع الويب بشكل مستمر. وتعمق برنامج الزاحف داخل موقع الويب وتحليل كامل صفحاته لأجل جمعها من شأنه رفع تكاليف العمل داخل محركات البحث. وتتجنب محركات البحث إظهار مدى العمق الذي تعمل به برامج الزحف داخل صفحات الويب.

- تأخر التحديث الدوري: فإضافة صفحة جديدة إلى محرك البحث يحتاج إلى وقت طويل قد يصل إلى شهر أو أكثر حتى يتم تمييزه ببرنامج الزاحف، وعلى الجانب الآخر فإن الرجوع إلى مواقع سابقة للبحث عن التحديثات الأخيرة بها من صفحات مضافة يحتاج إلى وقت أطول من الوقت الذي يستغرقه التحديث. وعلى ذلك فإن جزءاً من مواد المعلومات المحدثة داخل مواقع الويب التي تم إضافتها مسبقاً سيظل بعيداً عن أيدي الباحثين فترة من الوقت، وهو ما يحدث في قواعد البيانات كبيرة الحجم.

- ازدياد النتائج المستدعاة Maximum Number of Viewable Results : تتسم محركات البحث بكبر حجم قواعد بياناتها، ويتضح ذلك في أعداد النتائج المسترجعة داخل قاعدة البيانات التي قد يصل عددها إلى ملايين الصفحات، مما يعني معه أن محركات البحث تعمل على الإهمال المتعمد واستبعاد النتائج الأقل صلة بموضوع البحث، ومثل هذه النتائج تظل دائمًا الجزء غير المرئي أمام الباحثين.

- زيادة الروابط غير النشطة: وهذا بدوره لا يرجع إلى عمل محرك البحث وإنما خطأ في تفعيل روابط بعض الصفحات التي لا يستطيع برنامج الزاحف الوصول إليها إلا باستخدام مسار الصفحة URL أو الإضافة اليدوية عن طريق مصممي صفحات الويب.

١/٤/٢/٢ الويب الخاصة:

وهي تتكون من مجموعة من صفحات الويب ذات الطابع الشخصي سواء التي تتبع أفراد أو هيئات رسمية، ومن ثم لا توفر إمكانيات التفاعل مع محركات البحث . وعلى الرغم من قابلية هذه الصفحات للتكشيف بواسطة برنامج الزاحف، إلا أنها تستخدم حالة من الحالات الثلاث التالية لمنع دخول برنامج الزاحف إليها. وهذه الحالات هي:

- استخدام كلمات المرور للتعرف إلى محتوى الصفحة.
- تهيئة ملف robots.txt المحدد لعمل برنامج الزاحف لمنع التعامل مع صفحة الويب.
- استخدام حقل noindex meta tag للتوقف عن إضافة هذه الصفحة.

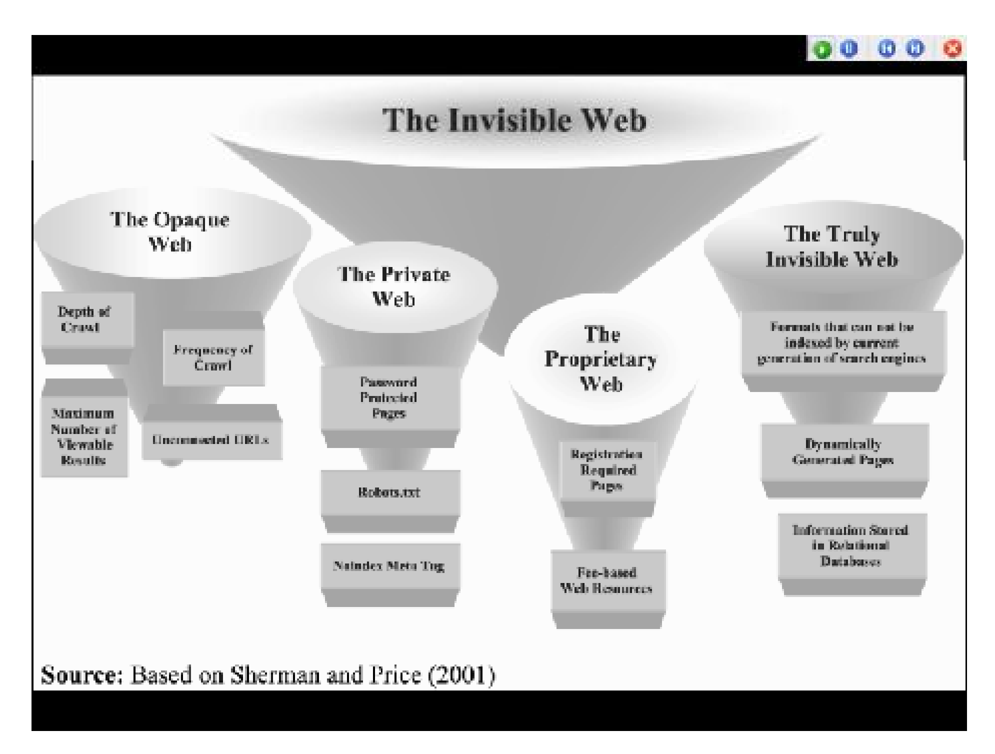
١/٤/٢/٣ الويب ذات الملكية:

ويقصد بهذا النوع مجموعة قواعد البيانات التي تعتمد على وجود تسجيل أو ربط Registration مع موقع القاعدة. وعلى الرغم من أن هذا النوع يقدم خدمات مجانية، فضلاً عن سهولة إتمام عملية الربط بالقاعدة، إلا أن أفضل برامج الزاحف حتى الآن لا تستطيع إقامة رابط لقواعد البيانات. مما يجعل من الصعب عبور الزاحف إلى محتوى هذه القواعد.

١/٤/٢/٤ الويب غير المرئية الحقيقية Thr Truly Invisible Web:

يعد هذا الجزء من الويب هو الواقع الفعلي لمفهوم الويب غير المرئية. ويمكن لأي جزء من صفحات الويب أن ينجلي أمام الباحثين إذا ما امتدت إليه برامج محركات البحث لتكشيفه وإخضاعه للبحث أمام المستفيدين، ويختلف الجزء

الحالي من الويب في البناء الفني عن الشكل المستخدم لدى برامج الزاحف، فلا يستطيع برنامج الزاحف استخدام محتوى تلك الصفحات. ويشير Paul Pedley إلى أن ذلك يتوقف على احتواء صفحات الويب على أنواع من الملفات لا تتعامل معها محركات البحث مثل الملفات المضغوطة وملفات pdf وملفات الفلاش Flash، ولا يكفي الجزء البسيط من حقول الميتاداتا المصاحب لهذه الملفات في إعانة برامج الزاحف على تحليلها وتكشيفها التي تحتاج إلى التعامل مع كامل المتن لصفحة الويب في تكشيفها وبحثها. ويوضح الشكل (١-٩) مكان الويب غير المرئية الحقيقية بين مختلف مستويات الويب غير المرئية الأخرى.



الشكل (١-٩) الويب غير المرئية الحقيقية بين مختلف مستويات الويب غير المرئية الأخرى[١]

(١) Ford, Nigel and Mansourian, Yazdan. The invisible web: an empirical study of "cognitive invisibility", Journal of Documentation,2006, cited at 9/12/2008,cited at http://www.emeraldinsight.com/10.1108/00220410610688732

وقد حددت Laura Cohen فئات عامة لمصادر الويب غير المرئية ، وهي التي تشكل الخصائص العامة لتكوين هذه الويب. وجاءت الفئات كما يلي[1]:

الفئة الأولى: محتوى قواعد البيانات:

يتكون محتوى قاعدة البيانات على الويب من معلومات تم اختزانها داخل جداول صممت باستخدام برامج مثل Access, Oracle, SQL Server and DB2 ، ويتم الوصول لهذا المعلومات باستخدام استفسارات البحث الملائمة لبنية قاعدة البيانات. وعلى ذلك فمحتوى قواعد البيانات يختلف عن محتوى صفحات الويب الذي يتم الوصول إليه مباشرة.

الفئة الثانية: ملفات الويب غير النصية :

مثل ملفات الوسائط المتعددة أو الرسومية أو البرامج أو تلك الملفات في شكل pdf , mdb, ppt and doc.

الفئة الثالثة : المحتوى محدد الوصول :

وهي مجموعة المواقع التي تم حمايتها باستخدام كلمات مرور أو برامج حماية؛ لطبيعة المحتوى الحساسة ، أو تلك الصفحات التي تتيح محتواها فقط من خلال استمارات الاشتراك لمستخدمي الويب.

الفئة الرابعة : المحتوى التفاعلي (الديناميكي) :

وهي صفحات الويب التي تضع استمارات إدخال للمعلومات وتتغير بياناتها بواسطة مستخدمي الموقع.

الفئة الخامسة : الصفحات فاقدة الروابط Unlinked content – pages :

حيث يسبب افتقاد هذه الصفحات لروابط فائقة مع الصفحات الأخرى قطع طريق الوصول أمام برامج الزاحف.

(١) Cohen, Laura. The Deep. internet tutorials, 2006, cited at 20/6/2007, cita at http://www.internettutorials.net/deepweb.html

الفئة السادسة : صفحات ذات تقنيات محددة Scripted content - pages :

حيث يتم الوصول لهذه الصفحات بواسطة روابط رسومية قد لا يتمكن برنامج الزاحف من التعامل معها.

١/٤/٢/٤/١ خصائص الويب غير المرئية:

إن مصطلح الويب غير المرئية قد نال اهتمامًا بالغًا في أدبيات الإنتاج الفكري حول بيئة الويب، وعلى الرغم من ذلك فإن بعض الباحثين مثل[١] Cohen, Laura يرون أن مصطلح "invisible Web" يعاني من فقر في الدلالة والتعبير عن واقع الويب. وذلك يرجع في نظرهم إلى الأسباب التالية:

١- يعبر مصطلح "invisible Web" فقط عن آلية عمل محركات البحث ؛ حيث يفترض المصطلح أن كل ما تصل إليه محركات البحث من مصادر الويب هو الويب المرئية، وأن غير المرئي من المصادر هو ما لا تستطيع الوصول إليه وتكشيفه داخل مستودعاتها. ومن ثم افتراض أن مستخدمي الويب يستطيعون فقط استرجاع معلومات الويب من خلال محرك البحث ، ولا أمل لهم في استرجاع معلومات أخرى خارج محرك البحث، إلا أن يلعب الحظ دوره في ذلك.

٢- لا يمكن القول إن هناك معلومات رقمية مسجلة يطلق عليها غير مرئية، وإنما يمكن القول أن هناك معلومات أصعب بدرجة ما في الحصول عليها من المعلومات الأخرى.

٣- إن قواعد البيانات على الإنترنت موجودة منذ وقت طويل، مما جعل مستخدمي الويب على قدر من التمرس بالتعامل وطرح الاستفسارات البحثية المباشرة لاسترجاع محتوى هذه القواعد.

---

(١) Cohen, Laura. The Deep.OP.cit.

٤- إن محتوى محركات البحث ذاته يتم اختزانه في قواعد بيانات، وهذا ألا يشكل بطبيعته جزءاً من الويب غير المرئية؟ حيث لا يستطيع مستخدم الويب استرجاع ما بداخل محرك البحث سوى باللجوء إلى محرك البحث وطرح الاستفسارات البحثية داخل واجهة البحث.

غير أن الباحث يتبنى رؤية مخالفة تمامًا ترد على الجوانب السابقة ؛ حيث تجاهلت الأسباب السابقة حقيقة أن محركات البحث هي واسطة العقد في عمل الويب بين منتج لصفحة الويب ومستخدم لها : فمن غير الممكن أبدًا أن يحفظ مستخدمو الويب جميع مسارات صفحات ومواقع وقواعد بيانات الويب. هذا إلى جانب أن عددًا قليلاً من محركات البحث يمكنه نظريًا تغطية كل المعلومات الرقمية الواقعة في فضاء الويب آخذين في الاعتبار التقدم والتطوير الذي تشهده محركات البحث.

إن ثمة حقيقة يجب التعرض لها وهي الربط بين قواعد البيانات عامة وبين الويب غير المرئية. فمصطلح الويب غير المرئية يطلق فقط على قواعد البيانات التي تعمل في بيئة الويب html, xml، لأن هناك الكثير من قواعد البيانات التي تعمل على الخط المباشر، والفارق جلي بين النوعين ؛ حيث تحمل قاعدة بيانات الويب خصائص خلاف ما توجد عليه قواعد الخط المباشر ويتضح ذلك من أساليب الاسترجاع والعمل لكل منهما. تختلف بيئة الويب عن الخط المباشر في ما يلي[١]:

- يعني مصطلح الخط المباشر تلك القواعد المتاحة على أقراص ضوئية CDs أو قاعدة محلية الاستخدام أو فهرس متاح على شبكة داخلية للمكتبة OPAC.
- ترتبط قواعد بيانات الخط المباشر بالتعامل مع الشكل العلائقي لقواعد البيانات RDBMS.

(١) online DB and searching.2004, cited at 15/7/2007, cited at http://web.simmons.edu/~benoit/LIS530/602slides/17online.htm

- يعني مصطلح الويب تلك الملفات المعالجة في بيئة html, xml ، التي يتم الوصول إليها عن بعد أو المختزنة في قواعد بيانات الويب.
- يعمل الاسترجاع على الخط المباشر بخصائص البحث والمضاهاة لمجموعات البيانات الببليوجرافية اعتمادًا على آليات البحث البولياني. أما الاسترجاع على الويب فغالبًا ما يعمد الى التكشيف والاسترجاع للنص الكامل، وهو ما تقوم به محركات بحث الويب.
- تعمل بيئة الويب كبيئة تفاعلية دائمة التغيير والإضافة والتحديث، خلاف ما يوجد عليه الخط المباشر من ثبات أو بطء في التغيير.

٥/١ خاتمة:

تحتوي الويب على كم هائل من المعلومات الرقمية تختلف في أشكالها وملفاتها ووسائط حملها. وتحفظ الويب هذه المعلومات في صفحة أو موقع أو قاعدة بيانات على الويب. وتتبنى الويب بناء مصادر معلوماتها بالاعتماد على لغات الترميز المعيارية التي تتوافق من ناحية أخرى مع آليات عمل أدوات البحث على الويب من محركات وأدلة بحث. . ولكل من صفحات ومواقع وقواعد البيانات خصائص محددة في المعالجة الفنية والتنظيم للمعلومات واليات محددة للاسترجاع. غير أن قواعد البيانات وبعض المصادر الأخرى قد شكلت بخصائص بنائها معلومات بعيدة عن متناول محركات البحث بما جعلها بعيدة عن متناول مستخدمي الويب.وعلى ذلك فأضحت الويب منقسمة بين محتوى مرئي لمحركات البحث ومستخدميها ، وبين محتوى آخر غير مرئي. ويتنوع الجانب غير المرئي من الويب بين الويب غير القابلة للاختراق أو الوصول من جانب برامج الزاحف، إلى تلك التي تحتاج لكلمات مرور ودخول شرعي إلى محتواها ، بجانب قواعد البيانات التي تتسم بكبر حجم معلوماتها بالدرجة التي تحتاج معها إلى مستوى أعمق عملية الزحف والوصول إلى أعمق صفحاتها وملفاتها.

الفصل الثاني

قواعد البيانات "الويب غير المرئية"

٢ / ٠ تمهيد:

تركز الدراسة في هذا الفصل على الشكل الأكثر أهمية من أشكال معلومات الويب غير المرئية وهو قواعد بيانات الويب. ويأتي محتوى قواعد بيانات الويب على رأس محتوى المواقع غير المرئية الأخرى كمًا وكيفًا. فتتميز قواعد البيانات كمصدر للمعلومات بطبيعة بناء وتنظيم خاص لملفاتها يعتمد في أساسه على بناء جداول البيانات. بالإضافة إلى خصائص أخرى مثل ؛ عمق مستويات التحليل واستخدام كلمات المرور ومحدودية الوصول للملفات. مما تسبب في جعلها الأكثر صعوبة في الوصول والمعالجة الفنية أمام أدوات تنظيم الويب. ويتناول هذا الفصل التعريف بقواعد بيانات الويب وكيفية تواجدها في بيئة الويب، انتقالاً إلى بناء قواعد البيانات على الويب، وآليات المعالجة والتنظيم والبحث بها ثم الحديث عن موقفها من برامج عمل محركات وأدلة البحث.

١/٢ قواعد بيانات الويب:

تتعدد جوانب الحديث عن قواعد بيانات الويب؛ حيث يمكن تناولها من ناحية الوصول والتنظيم والمعاجلة لملفاتها، ثم تكاملها مع برامج تصفح الإنترنت والآليات المستخدمة لتحقيق التوافق بينهما. ويشمل الحديث عن قواعد بيانات الويب الجوانب التالية:

١/١/٢ ماهية قواعد بيانات الويب:

تعتمد بنية قواعد البيانات على شكل ثابت سواء أكان ذلك في بيئة الويب أم خارجها. وقد اتفقت تعريفات قواعد البيانات خارج بيئة الويب مع التعريفات التي تخص قواعد بيانات الويب في جانبي البناء والتنظيم، وقد اختلفت التعريفات فقط في جانب التوافق بين بيئة html وبنية قواعد البيانات عند الحديث عن قواعد بيانات الويب. ومن أهم التعريفات مثل "قاعدة البيانات هي نظام تم

تصميمه باستخدام آليات محددة تقوم بتنظيم وحفظ واسترجاع المعلومات، وتعتمد في أداء ذلك على بناء جداول المعلومات." غير أن هذا التعريف يعد بسيطا إلى حد التعريف العام بقاعدة البيانات لأنه لم يتطرق إلى العلاقة بين الويب وقواعد البيانات"[١].

وهناك تعريف آخر لقاعدة بيانات الويب على أنها " قائمة منظمة لصفحات الويب. وهي تشبه إلى حد كبير الفهرس البطاقي للمكتبات ؛ حيث تعتمد قاعدة البيانات على وصف صفحات الويب من خلال عناصر مختارة مثل العنوان والموضوع. وهذه العملية تسمى بالتكشيف، وقد تحتوي قاعدة البيانات على ملايين من صفحات الويب بهذه الطريقة." غير أن المتمعن في هذا التعريف يرى أنه تعريف لمحركات البحث وليس لقاعدة البيانات ؛ حيث إن محركات البحث تعرف خطأ بقواعد بيانات الويب لما لها من وظيفة اقتناء وتنظيم وبحث مواقع الويب[٢].

وقد وصفت بعض التعريفات الأخرى قواعد بيانات الويب بما يعكس واقعها الفعلي على الويب، على أنها " وعاء للمعلومات ذات بنية وتصميم محددين. ويتم اختزان المعلومات بها ومعالجتها ثم استدعاؤها بوسائط حمل متنوعة، وتستخدم لأغراض متعددة. ويستعين مستخدمو قاعدة البيانات في ذلك باستراتيجيات بحث مختلفة." وقد تضمن التعريف هنا مفهوما آخر يختص باستخدام قاعدة البيانات عن بعد، وحين يتم وضعها على حاسب مركزي central computer أو خادم إنترنت internet server يمكن الوصول إليها بإحدى الطرائق التالية:

(١) Chapple , Mike. What is a web database. About, Inc, 2007, cited at 15/6/2007, cited at http://databases.about.com/od/specificproducts/

(٢) Introduction to Web Databases. 2002 by Scott Nicholson. Cited at http://www.askscott.com/sec2.html

أولاً: تعتمـد قاعـدة البيانـات علـى التكامـل مـع بنيـة HTML في عـرض مجموعاتها ونتـائج البحـث أمـام المستخدمين في صفحات موقـع قاعـدة البيانـات. وذلك كما هو الحال في فهارس الخط المباشر OPAC .

ثانيًا: تستخدم قاعدة بيانات الويب نماذج عرض وقوالب ثابتة تعرض معلوماتها من خـلال صفحة الويب. ويسمى موقـع الويب في هـذه الحالـة بموقـع ديناميكي dynamic site، كما يطلق علـى قاعـدة البيانـات هنـا مصطلح نظـم إدارة المحتوى content management system or CMS.[1].

يخلص الباحث من التعريفات السابقة إلى رؤية أكثر تحديدا لقواعد بيانـات الويب على أنها "كم هائل مـن المعلومـات يتم بنـاؤه وتنظيمـه باستخدام نظـم إدارة قواعـد البيانـات مثـل MsAccess, Oracle, SQL, MySQL لأداء مجموعـة مـن الوظـائف كالإضـافة والتنظيم والبحـث. وتـؤدي قاعـدة البيانـات عملـها في بيئـة الويب مستعينة ببرامج تحويل بين بيئة الويب HTML ونظام إدارة قاعدة البيانات" . وتهدف قاعدة بيانات الويب إلى أداء مهام محددة في معالجة وتنظيم البيانات هـي:

١- المعالجة والتنظيم لمجموعات المعلومات التي تختزنها.

٢- القيـام بمهـام الإضـافة والحـذف والتعـديل والتحـديث لمعلومـات قاعـدة البيانات.

٣- تجهيز المعلومات وعرضها وفقا للحاجات الموضوعية لمستخدمي الويب.

٢/١/٢ قواعد البيانات كجزء من الويب:

تهدف مواقع الويب التي تتيح قواعد بياناتها على الإنترنت إلى إيصال محتواهـا لكل مستخدمي الويب على اختلاف أماكنهم، وتستخدم لـذلك تقنيـة الاتصال بالإنترنت لتحقيق وتقديم وتوصيل ملفات قواعد البيانات إلى مستخدميها. ومـن ثـم تمثـل الويـب هنـا مجـرد ثـوب جديـد ترتديـه قواعـد البيانـات لتـتلاءم مـع طبيعـة

---

(١) Web Databases. MDA, 2002,cited 25/6/2007, cited at http://www.mda.org.uk/webdatabases.htm

الإنترنت مكانًا وزمانًا. وتختلف قواعد البيانات في درجة ملاءمتها مع بيئة الويب ؛ حيث تعد بنية قواعد البيانات العلائقية relational databases أكثر النماذج كفاءة في تنظيم وبحث المعلومات.

وتعتمد قواعد البيانات العلائقية على وسيلة المفاتيح العامة أو الرئيسية common key للربط بين ملفات قاعدة البيانات بعضها ببعض. ودائمًا ما يعطى هذا المفتاح للحقول أو العناصر التي لا يتكرر محتواها داخل تسجيلات قاعدة البيانات مثل حقول ID or reference number.

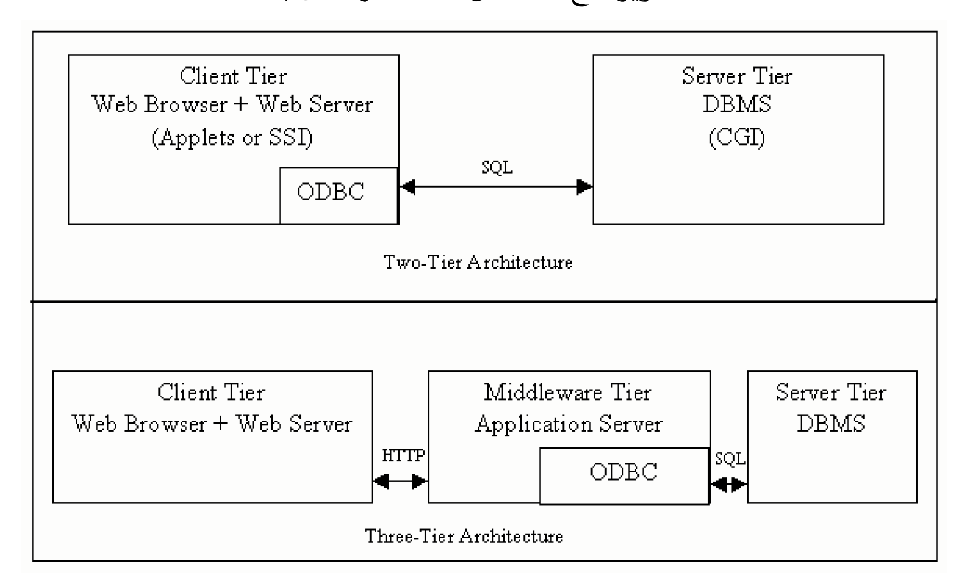
تبعد المسافة بين قاعدة البيانات والمستخدم في بيئة الويب عنها في البيئة الرقمية الحاسوبية ؛ حيث يتوسط خادم الويب العلاقة بين قاعدة البيانات ومستخدمي القاعدة. ودور الوسيط يلزم خادم الويب بحمل سؤال البحث search questions من المستخدم إلى قاعدة البيانات ثم استقبال النتائج search results وإخراجها إلى المستخدم بما يوافق شاشات متصفحات الويب.

وقد يساعد على الأداء المثالي لقاعدة البيانات على الويب الاختيار الجيد لنظام إدارة قاعدة البيانات ؛ فيعتبر نظام Structured Query Language (SQL) من أفضل نظم إدارة قواعد البيانات العلائقية حيث يمكن للمستخدم التعامل معه دون الحاجة لمستوى مهارة أو خبرة عالية بطبيعة عمل قواعد البيانات العلائقية. ويتميز هذا النظام أيضًا بالقدرة على التعامل مع قواعد البيانات كبيرة الحجم، غير أنه يحتاج إلى نفقات عالية تصل إلى آلاف الدولارات[1].

وثمة جانب آخر يؤخذ في الاعتبار وهو مدى توافق نظام إدارة قواعد البيانات التي تعمل في بيئة الويب مع برامج التحويل أو إنتاج واجهات عرض قاعدة البيانات. وتعد برامج PHP or perl programming هي الأكثر توافقًا للعمل مع نظام إدارة قواعد البيانات MySQL.

(١) Web Databases. MDA.Lbid.

تكـون قواعـد بيانـات الويـب مـع خادمـات ومسـتخدمي الويـب علاقـة ثنائيـة أو ثلاثيـة الترابـط، ويـتم التعـبير عنهـا بالمصطلحات Two-tier Architecture OR Three-tier Architecture. ويوضح الشكل (٢-١) أركان بناء هذه العلاقة :



الشكل (٢-١) العلاقة ثنائية أو ثلاثية الترابط لقواعد بيانات الويب

يوضح الشكل (٢-١) الجوانب التالية:

١- ترتبط قواعد بيانات الويب مع مستخدمي الويب بطريقتين أساسيتين ؛ إما طريقة ثلاثية البناء أو ثنائية البناء.

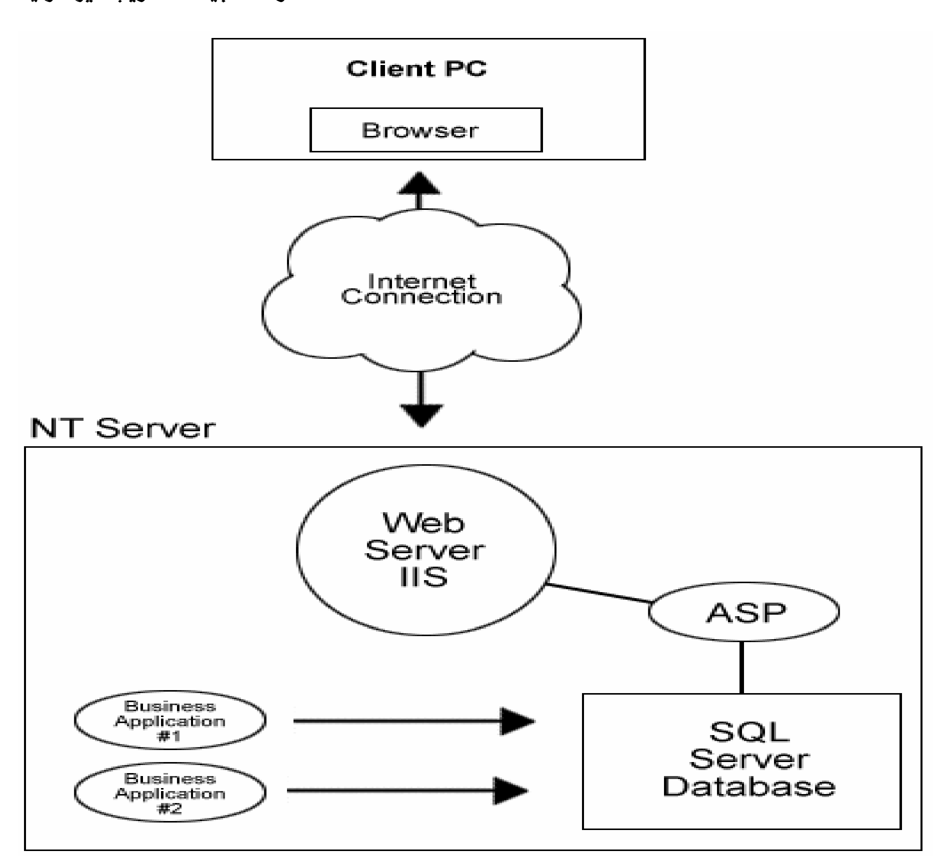
٢- تستخدم لغـات التحويـل CGI في الطريقـة الثنائيـة في جانـب قاعـدة البيانـات لتحويـل الـوارد إليهـا مـن الويـب إلى لغـة نظـام قاعـدة البيانـات، وتـستخدم معدل الجافا Java applets أو SSI في جانب خادم الويب.

٣- تختلف الطريقة الثلاثية عن الثنائية لعمل قواعد بيانـات الويـب في احتوائهـا علـى وسـيط بـين خـادم قاعـدة البيانـات وخـادم الويـب متمـثلا في خـادم التطبيقـات، المسئول عـن نقـل إشـارات المعلومـات بلغتيهـا بـين خـادم قاعـدة البيانات وخادم الويب.

٣/١/٢ أداء قواعد البيانات في بيئة الويب:

تعتبر قاعدة بيانات الويب الضلع الثالث في مثلث نظام استرجاع المعلومات على الخط المباشر، حيث يتكامل معها مستخدم الويب client PC من ناحية وخادم الويب web server من ناحية أخرى. وقد أشارت التعريفات السابقة إلى أهمية التوافق بين نظام إدارة قاعدة البيانات وبين بيئة الويب لتحقيق الانسجام في عرض المعلومات أمام مستخدم الويب. ومن ثم تكون الويب البيئة الرقمية التي تتيح قاعدة البيانات للاستخدام أمام المستفيدين. وأصبح أمام الويب الحاجة الدائمة إلى تطوير التوافق بين قواعد البيانات وخادمات الويب باستخدام آليات ASP, PHP حتى تنساب المعلومات بين قاعدة البيانات ومستخدميها. ومن بين الشركات التي حرصت على ذلك شركة مايكروسوفت Microsoft.inc حيث قدمت شكلاً مقترحًا لتحقيق التوافق المرجو من استخدام قواعد بيانات الويب. ويرى الباحث أن إخضاع قواعد البيانات ببنيتها ونظم إدارتها المختلفة يبدأ من إذابة الحدود الرقمية بين بنية الويب وأدوات بحثها من جانب، وبين بنية قواعد البيانات من جانب آخر. وذلك يمكن تحقيقه في حالة الاهتمام ببرامج التحويل أو واجهات الدخول بين خادمات الويب ونظم إدارة قواعد البيانات، كما يتضح في الشكل (٢-١) التخطيط العام للربط بين قواعد البيانات وخادمات الويب[1]:

(١) Summary of a New Client/Server/Database Architecture.2003, cited at 25/6/2007, cited at http://www.semibase.com/database.html



الشكل (٢-٢) التوافق بين مستخدم الويب والخادم وبنية قاعدة البيانات

يتضح من الشكل (٢-٢) النقاط التالية:

- أن الأداء المثالي لقاعدة البيانات على الويب يعتمد على أركان ثلاثة، يفشل الأداء إذا ما غاب أحدها.
- أن برامج التحويل scripting language هي الداعم الأساسي في تطويع قواعد البيانات للعمل في بيئة الويب.
- تستطيع أدوات بحث الويب استخدام برامج التحويل لتطوير برامج الزاحف لتخطي حاجز الويب إلى قاعدة البيانات، ومن ثم الدخول إلى محتواها.

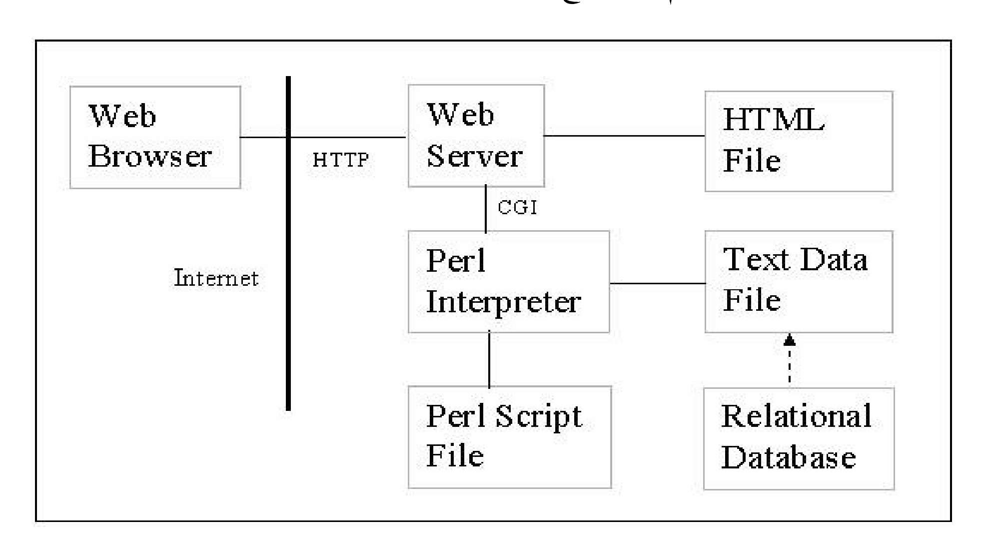
وهناك برامج التحويل المتعددة التي تساعد على تطويع محتوى قواعد البيانات للعرض في واجهات الويب، ومن أهمها البرامج التالية:

- **Common Gateway Interface (CGI)**
- **PHP**
- **Active Server Pages (ASP) , VBScript and ActiveX**
- **Java / JDBC / JSQL**
- **Javascript / JScript**
- **Server-Side Directives**
- **Cookies**
- **Server Application Programming Interface (API)**
- **ColdFusion**
- **CORBA**

غير أن البرنامج الأشهر في الاستخدام من بين هذه البرامج هو Common Gateway Interface (CGI) وتعتمد إصدارة CGI-Based System Using Perl على الوصول المباشر أو غير المباشر إلى قاعدة البيانات، ومن ثم تقوم بإنتاج شاشات أو ملفات نصية لعرض ملفات قاعدة البيانات. غير أن برنامج CGI قد يصاب بضعف الأداء داخل خادمات الويب نتيجة لكثرة العمليات المستمرة أو تنفيذها، أو تغيير نمط الإتاحة من مستخدم لآخر. كما أنها تضعف باستمرار ديناميكية قاعدة البيانات في إعادة بناء ملفاتها.

وتشكل مجموعات المعلومات كبيرة الحجم عبئًا آخر على برنامج CGI، حيث يتطلب ذلك شغل حيز كبير من عمل البرنامج، بالإضافة إلى متطلبات تحويل وصياغة استفسارات البحث الموجهة إلى قاعدة البيانات. ويزداد كل ذلك صعوبة إذا ما علمنا أن برنامج CGI يؤدي عمله من إرسال واستقبال هذه العمليات المعقدة مع قاعدة بيانات علائقية توجد على حاسب أو خادم آخر، بالإضافة إلى أن مختلف إجراءات البرنامج تتم بالاعتماد على بيانات تعليمات ملف نصي.

وإلى جانب ما سبق من أوجه ضعف فإن برنامج CGI-Based System Using Perl أيضًا لا يساعد على أداء عمليات التحديث المباشر لقواعد البيانات من خلال بيئة الويب. وعلى ذلك فإن هذا البرنامج يحتاج دائمًا إلى التطوير الملائم للتعامل مع بيئة الويب التي تستخدم تقنيات رسومية JAVA Scripts متطورة. ويوضح الشكل (٢-٣) عمل CGI في نظام استرجاع محتوى قواعد البيانات على الويب(١).



الشكل (٢-٣) عمل CGI في نظام استرجاع محتوى قواعد البيانات على الويب

وقد ذكرت بعض الدراسات عن قواعد بيانات الويب Web-based Databases سلبيات وإيجابيات استخدام برنامج CGI ، وهي كالتالي(٢):

أولاً: الإيجابيات :

- استقلالية خادم الويب: لا تعتمد في أدائها على خادمات الويب بما يمكنها من العمل بمرونة عالية.

(١) Lazar, Z. Peter and Holfelder, Peter. Web database connectivity with scripting languages O'Reilly & Associates, 1997, cited at 25/6/2007, cited at http://www.oreilly.com/catalog/wj6/

(٢) Web-based Databases: technical Implementation.ppt .cited 30/6/2007, cited at http://foodafrica.nri.org/enterprises/enterprisespapers/GhirmaMogesQuality.doc

- استقلالية اللغة : تستجيب للتعامل مع تيجان Tags مختلفة من اللغات.
- التفضيل والقابلية بين مصممي الويب.
- البساطة والسهولة في الاستخدام.
- البرنامج الأقدم والأكثر انتشارًا ومعرفة بين مستخدمي ومصممي الويب.

ثانيًا: السلبيات :

- إن مختلف إجراءات وعمليات CGI التي تتم بين المستخدم وقاعدة البيانات، لابد أن تمر من خلال خادم الويب، حيث يصوغ خادم الويب استفسارات مستخدم الويب وينقلها إلى قاعدة البيانات، كما أن عليه أيضًا ترتيب وإعادة ترتيبها على شاشة العرض. مما قد يحدث خطأ في التعامل مع إجراءات ذات بنية HTML .
- تعاني برامج CGI من نقص الحماية والأمن ضد الاختراقات (backdoors).
- تسبب الإجراءات والعمليات المتواصلة لـ CGI ضغطًا كبيرًا على قدرة عمل خادم الويب.
- يفتقد برنامج CGI إلى الفاعلية الكافية للتعامل مع أسئلة البحث التفاعلية، أي تلك الأشكال المختلفة من سؤال بحث مستخدم محدد أو أسئلة بحث مستخدمين متعددين. كما يصعب عليها أيضًا ترجمة الأشكال المختلفة لأسئلة البحث التي تطرحها قاعدة بيانات الويب أمام مستخدميها.

١/٣/١/٢ قواعد البيانات بين اللغة المعيارية HTML والقابلة للتوسع XML:

تعتبر لغة الترميز المعيارية الممتدة XML الشريك الثاني للغة HTML في بناء واجهات الويب لقواعد البيانات والمؤثر في عملها. ويمكن القول إن هناك خصائص فارقة بين اللغتين تمكن اللغة الممتدة من أن يكون لها الصلاحية

الكبرى في تبادل وتداول ملفات وصفحات الويب. وتمثل الخصائص بالشكل التالي:

- لم يكن في ذهن مصممي لغة الترميز HTML تبادل ملفات الويب للأغراض التجارية، بقدر ما كان ذلك في تصميم وإنشاء لغة الترميز XML.
- لقد صممت لغة الترميز HTML بالاعتماد على تيجان لغة محددة encoded LANGUAGE. حيث لا يمكن على سبيل المثال تبادل وثيقة رقمية بين أيرلندا باللغة الإنجليزية وكوريا الجنوبية باللغة الكورية (Hangul). بخلاف الترميز الموحد Unicode الذي تعتمد عليه لغة XML؛ مما يمكنها من تبادل الملفات بين أنظمة الحاسبات المختلفة، بالإضافة إلى تبادل الملفات بين الثقافات المختلفة.

٢/٢ أشكال قواعد البيانات على الويب :

إن تواجد قواعد البيانات على الويب لا يأخذ فقط شكل قاعدة البيانات التي تنظم وتبحث ملفات بشكل خاص في موضوع خاص، وإنما تتعدى ذلك إلى التواجد بأشكال مختلفة تتحد وفقاً لطبيعة البناء ونوع الرسالة لموقع الويب. بل يمكن القول إن مصادر معلومات الويب مرت بمراحل تطور مختلفة منذ بدأت بمحتوى ثابت لصفحة HTML واحدة، حتى وصلت إلى قواعد بيانات النصوص الكاملة. وهي بذلك تنتقل خطوة تلو الأخرى باتجاه التحول من الشكل الثابت للمحتوى static and file-based إلى الشكل التفاعلي في طرح أسئلة البحث وتقديم النتائج Dynamic Web Resources. وقد ذكرت بعض الدراسات في تقديمها لقواعد البيانات هذه المراحل وبعض خصائصها كالتالي[(١)]:

(١) Multimedia Database Management Systems.ppt Cited at 30/6/2007, cited at http://www.chu.edu.tw/~leecc/962/B10204B/1%20Databases%20and%20Database%20 Management%20Systems.ppt

أولاً: مرحلة المحتوى الثابت static contenet :

- لقد حملت هذه المرحلة سلبيات في الأداء مثل الإسهاب في الإنتاج والتكرار في المحتوى وغياب النماذج المعيارية إضافة إلى ضعف الاستخدام المتعدد multi-user.
- يظل المحتوى الثابت على حاله دون تغيير أو تعديل حتى يتغير ملف المعلومات ذاته بملف آخر.

ثانيًا: مرحلة الويب التفاعلية Dynamic Web:

وهي تتسم بالخصائص التالية :

- الاستجابة لمدخلات مستخدم الويب ببعض الأشكال مثل نماذج إدخال البيانات HTML form.
- الاستجابة لرغبات مستخدمي الويب customize من خلال التعامل مع بطاقات التعرف إلى مصادر الويب sites cookies.

ثالثًا: مرحلة التحول إلى قواعد البيانات:

- عند بلوغ مستخدم الويب لموقع وإن كان ذا عدد محدود من صفحات الويب، فإن حاسبه الشخصي يعد كخادم الويب لهذا الموقع.
- يتطلب موقع تشغيل قاعدة البيانات لمساحة كبيرة من ذاكرة العمل لخادم الويب، مع توافر القدرة على معالجة عمليات متعددة في وقت واحد multiprocessing capabilities.
- تعد قواعد البيانات بمثابة العمود الفقري للتجارة الإلكترونية والتسويق الرقمي.

رابعًا: مرحلة ما بعد قواعد البيانات:

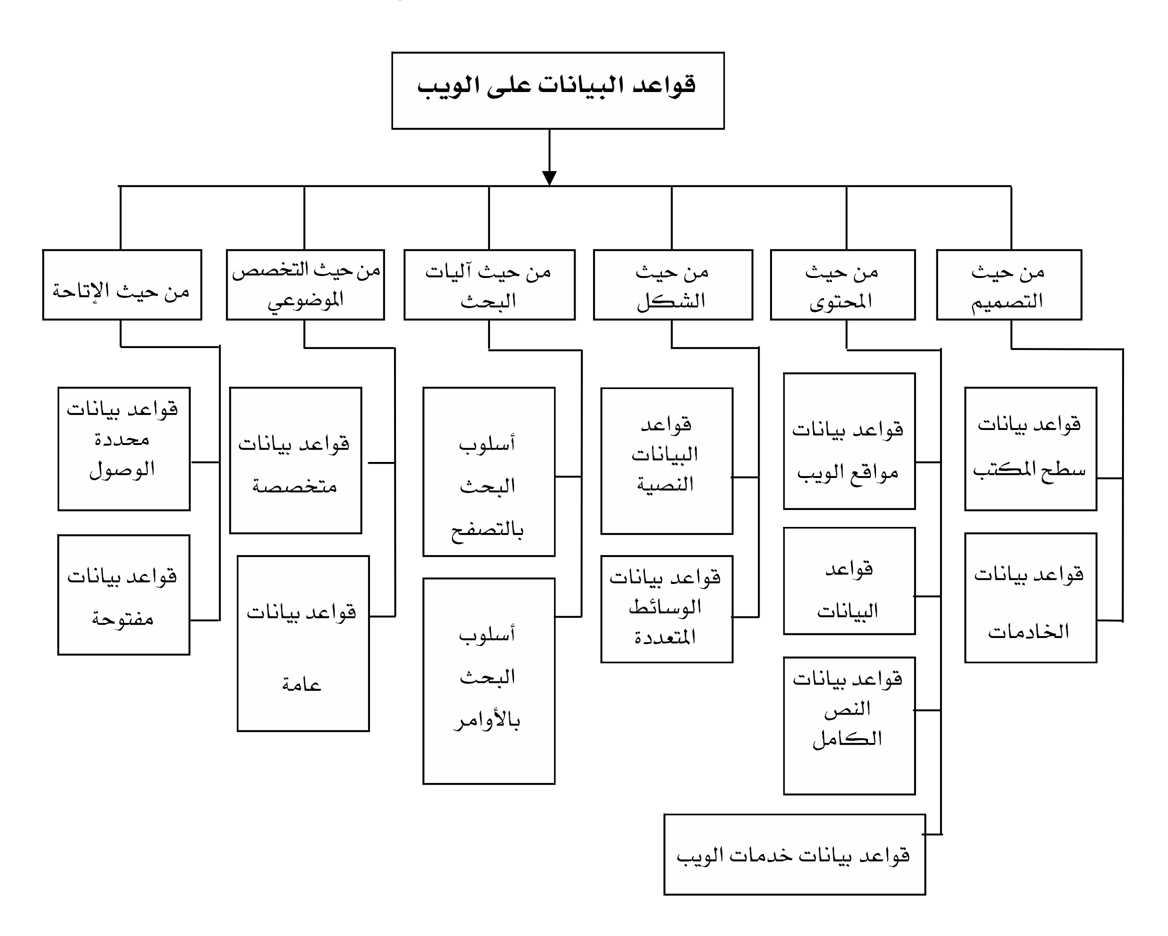
- وأهم ما يذكر عنها، زيادة التفاعلية بين مستخدم الويب والمحتوى الرقمي؛ حيث يستطيع مستخدم الويب التغيير والتعديل في سمات المحتوى الرقمي، وليس فقط عرضه أو استخدامه.

- القدرة الكبيرة لبعض المواقع مثل أدوات البحث في إنتاج محتوى متضمن لمحتوى مصادر الويب مثل التنويهات والشعارات والدعاية التجارية.

لقد مكنت القدرات التشغيلية والتخزينية العالية لخادمات الويب المستخدمين من بناء قواعد بيانات تستطيع خدمة مواقع الويب. وتنوعت أشكال المحتوى داخل قواعد بيانات تلك المواقع حيث تختزن بعض قواعد البيانات بيانات العملاء لمواقع التجارة الإلكترونية، إلى جانب حفظ وإدارة التسجيلات الببليوجرافية لمواقع المكتبات ونظم استرجاع المعلومات، فضلاً عن قواعد بيانات الهيئات الحكومية وتقديم الخدمات... وغيرها من الأشكال الخدمية التي توجد عليها قواعد البيانات، وهو ما يرجع بدوره إلى طبيعة وهدف إنشاء موقع الويب.

زادت أعداد قواعد بيانات الويب إلى الدرجة التي أصبحت تشكل فيها الجزء الأكبر من مصادر معلومات الويب. ونتيجة لذلك تنوعت قواعد البيانات شكلاً وبنية ونوعًا؛ مما أوجد معه اكثر من أساس تقسم عليه قواعد البيانات. وقد اقترح الباحث التقسيم التالي لتوضيح أشكال قواعد البيانات على الويب، يعرضها الشكل (٢-٤):

الشكل (٢-٤) أشكال قواعد البيانات على الويب



١/٢/٢ من حيث التصميم:

يوجد الكثير من برامج تشغيل قواعد البيانات تختلف بين البرامج البسيطة والمعقدة مثل ؛ Oracle, SQL Server, Microsoft Access, MySQL, DB2, Paradox، غير أنها جميعًا تندرج أسفل تقسيم عام إلى برامج قواعد بيانات سطح المكتب Desktop Database Systems ، وبرامج تشغيل قواعد بيانات الخادمات Server Database Systems. يتم اعتماد أي من البرامج السابقة وفقًا لاحتياجات العمل

ولطبيعـة البيانـات والأداء المتوقـع مـن قاعـدة البيانـات، وعلـى ذلـك فـإن تحليـل الاحتياجات هو المحدد لنوع قاعدة البيانات قبل الشروع في إنشائها، وتتم عملية التحليل اعتمادًا على أسئلة مثل[1]:

- ما الكيفية التي ستستخدم بها قاعدة البيانات، والمهام التي ستقوم بها؟
- ما الكيفية التي ستضاف وتعدل بها البيانات؟ ومن الذي سيقوم بذلك؟
- من الذي سيدعم جوانب تكنولوجيا المعلومات في عمل قاعدة البيانات؟
- ما الأجهزة المتاحة لعمل قاعدة البيانات؟ وهل توجد ميزانية كافية لإضافة بعض الأجهزة الأخرى؟
- من سيكون المسئول عن صيانة البيانات في قاعدة البيانات؟
- هل ستتاح قاعدة البيانات على الويب، وإذا كان ذلك فما مستوى الإتاحة للبيانات؟

عند الإجابة عن الأسئلة السابقة يمكن للقائمين على التحليل تقرير أحد نوعي قواعد البيانات حسب نوعية التشغيل كالتالي:

١/١/٢/٢ قواعد بيانات سطح المكتب Desktop databases:

تقدم قواعد بيانات سطح المكتب حلاً بسيطًا وغير مكلف لتلبية الاحتياجات البسيطة في اختزان كم قليل من البيانات ومعالجة غير معقدة لتنظيم البيانات. ومما يدل على بساطة العمل بهذا النوع اقتران قواعد البيانات بمصطلح Desktop الذي يرمز للحسابات الشخصية PCs. ويتسم هذا النوع بملامح مثل:

---

(١) Chapple , Mike. Choosing a database for your organization.2005, cited at 10/10/2008, cited at http://databases.about.com/od/administration/a/choosing_a_db.htm

- يعد هذا النوع غير مكلف ماديًا من قواعد البيانات ؛ فهو لا يحتاج سوى الحاسبات الشخصية. مقابل إنفاق آلاف الدولارات لقواعد بيانات الخادمات.
- يتسم هذا النوع بسهولة وقابلية الاستخدام عن قواعد بيانات الخادمات التي تحتاج لمهارات تدريب أكثر صعوبة.
- يصمم هذا النوع للعمل من جانب مستخدم واحد في أغلب الأحيان.
- تقدم أغلب قواعد هذا النوع القابلية والتوافق للعمل في بيئة الويب، وهو ما يسهل إتاحتها على الخط المباشر.

ومن أشهر قواعد البيانات لهذا النوع: Microsoft Access, FoxPro, FileMaker Pro, Paradox and Lotus Approach

٢/١/٢/٢ قواعد بيانات الخادمات:

يقدم هذا النوع فرصًا أكبر لتخزين ومعالجة كم هائل من المعلومات بعكس النوع السابق. كما يتيح إمكانية تشغيل قاعد البيانات من خلال أكثر من مستخدم في وقت واحد. ويحقق هذا النوع كثيرًا من الفوائد لتشغيل قاعدة البيانات، مثل:

- **قدرات العمل الفائقة:** حيث يمكن لهذا النوع العمل مع أكثر من مستخدم وأداء مختلف التطبيقات التي يتطلبها العمل.
- **المرونة :** يستطيع هذا النوع العمل في بيئات مختلفة لما يملك من إمكانات إنتاج وتصميم الواجهات application programmer interfaces (or APIs) حيث يمكن لقاعدة البيانات Oracle platform العمل وفقًا لنظم تشغيل مختلفة. وهذه النقطة مهمة في إطار العمل في بيئة الويب.

- **الأداء القوي:** يملك هذا النوع معالجات للعمل تستطيع التوافق مع خادمات العمل ذات الأجهزة المتطورة بما يحقق أداء سريعًا فعالاً لعمليات تشغيل قاعدة البيانات.
- **التوافق :** وهذا العنصر يرتبط بالسابق ؛ حيث يمكن لقاعدة البيانات تحقيق التوافق مع الأجهزة المتطورة والعمل بكفاءة عالية.

ومن أشهر قواعد البيانات لهذا النوع، قواعد: Microsoft SQL Server, Oracle and IBM DB.

٢/٢/٢ من حيث المحتوى:

تختلف قواعد بيانات الويب بحسب ما تحويه من معلومات، فبحسب ما تحويه قاعدة البيانات من أشكال المعلومات تأخذ المسمى المحدد لهذا المحتوى. وأفضل ما يدلل على ذلك أن بعض الباحثين يطلقون مسمى قواعد بيانات الويب على محركات بحث الويب، انطلاقًا مما تحويه مستودعات محركات البحث من صفحات ومواقع الويب. وتحوي قواعد البيانات أيضًا بيانات وسمات العملاء للمواقع التجارية، وأسماء وبيانات المشتركين لمواقع خدمات الويب مثل مواقع البريد الإلكتروني، وتحوي البيانات الببليوجرافية والمستخلصات ككشافات للدوريات الإلكترونية، وتحوي النصوص الكاملة كما في الصحف الرقمية والمواقع الإعلامية بجانب قواعد بيانات النصوص الكاملة في المجالات الموضوعية المتخصصة. وانطلاقًا من محتوى قواعد بيانات الويب فإنها تقسم إلى الفئات الأربعة التالية:

١/٢/٢/٢ قواعد بيانات مواقع الويب (أدوات البحث):

تعمل مستودعات أدوات بحث الويب على اختزان وتنظيم كل ما يمكن أن تصل إليه من صفحات ومواقع وملفات الويب، ولذلك كان لزامًا أن يطلق عليها قواعد بيانات الويب ؛ حيث يمكن القول – نظريًا – إن أدوات بحث الويب من

محركات وأدلة وبوابات إنما تحوي كل التراث الرقمي لبيئة الويب من صفحات ومواقع. ويطلق عليها من ناحية أخرى فهارس الويب أو كشافات الويب نظرا لما تؤديه من دور في تنظيم ومعالجة مصادر الويب. تتلقى قواعد محركات البحث كل ما يرد إليها من صفحات ومواقع يقوم برنامج الزاحف بنسخها وضمها إلى قاعدة البيانات. وعلى ذلك فلا يمكن تخيل الويب بدون تلك القواعد التي تعمل كجسر يمتد من مصادر الويب إلى مستخدمي الويب لتحقيق ما يحتاجونه من معلومات رقمية[1].

ونماذج قواعد بيانات الويب لا حصر لها، غير أن أهمها :

- **Google** : **www.google.com**
- **MSN** : **www.msn.com**
- **YAHOO** : **www.yahoo.com**

٢/٢/٢/٢ قواعد البيانات الببليوجرافية:

نقدم قواعد البيانات الببليوجرافية المعلومات التي تساعد في حصر وضبط واسترجاع الإنتاج الفكري. ويأتي دور هذه البيانات في تمثيل مصادر المعلومات تمثيلاً مكثفًا؛ حيث يتم التعويض عن مصدر المعلومات كاملاً بمجموعة من الحقول الواصفة أو العناصر التي تحكي صفة هذا المصدر وموضوعه. وتتخذ قواعد البيانات الببليوجرافية أشكالاً عدة، فهناك قواعد البيانات الببليوجرافية الحاصرة مثل فهارس المكتبات على الخط المباشر OPAC وقواعد بيانات المستخلصات وقواعد بيانات الكشافات التي تشتمل على المصطلحات الكشفية المعبرة عن محتوى مصدر المعلومات.

ومن نماذج هذه القواعد على الويب :

---

(١) Website databases and programming. DianeV. Web design studio, 2007,cited at 20/6/2007, cited at http://dianev.com/internet-marketing.html

- الفهرس الرقمي لمكتبة مبارك العامة : mpl [١].
- مستخلصات الرسائل الجامعية : UMI proquest [٢].
- كشاف الدوريات الطبية : Medline [٣].

٣/٢/٢/٢ قواعد بيانات النص الكامل:

تقدم قواعد بيانات النص الكامل على الويب لمستخدم الويب المعلومات النهائية لموضوع بحثه. وهي تتعدى بذلك مستوى التوثيق والحصر الذي تقف عنده قواعد البيانات الببليوجرافية. ونتيجة للقدرات التقنية التي تشهدها الويب سواء في التخزين والحفظ أو في المعالجة والبحث، فإن قواعد البيانات تستطيع تشغيل كم هائل من النصوص الرقمية بجانب وسائط حمل المعلومات الأخرى. وتختلف درجة الإتاحة بين تلك القواعد بين ما هو متاح وممنوع ومدفوع، وذلك إنما يتوقف بالدرجة الأولى على طبيعة موقع الويب المتضمن لقاعدة بيانات النصوص الكاملة.

ومن نماذج هذه القواعد :

- قاعدة بيانات find articles[٤].

٤/٢/٢/٢ قواعد بيانات خدمات الويب:

تقدم مواقع الويب على اختلافها خدمات متنوعة تسعى بها إلى اجتذاب أكبر عدد من مستخدمي الويب على اختلاف فئاتهم العمرية. وبحسب طبيعة خدمات موقع الويب يتحدد محتوى قاعدة بياناته. ولا يقتصر الأمر هنا على خدمات الترفيه أو الخدمات العامة التي تقدمها مواقع الويب، وإنما يمتد الأمر ليشمل المواقع الهيئات التجارية والحكومية التي تقدم خدماتها لمواطني الدولة التي تنتمي

(١) http://catalog.mpl.org.eg/
(٢) http://wwwlib.umi.com/dissertations
(٣) medline.cos.com/http://
(٤) www.findarticles.com

لها ، وتؤدي هذه المواقع وظائفها اعتمادا على قواعد بيانات تحمل نماذج العمل وبيانات العملاء وتقارير المهام...وغيرها.

وأشهر قواعد بينات خدمات الويب ؛ خدمة البريد الإلكتروني Hotmail. كما أن موقع الحكومة المصرية يعد نموذجًا لقواعد بيانات الهيئات الحكومية (موقع الحكومة المصرية)[1].

٣/٢/٢ من حيث الشكل:

تملك الويب أشكالاً مختلفة من المعلومات الرقمية تخدم مختلف الأغراض البحثية لمستخدمي الويب. وتبعًا لأن قواعد بيانات الويب هي شريحة من شرائح مصادر المعلومات الويب، فإنها تحوي مختلف أشكال ووسائط حمل المعلومات الرقمية. وتنقسم المعلومات الرقمية عامة داخل قواعد البيانات إلى شكلين، هما:

١/٣/٢/٢ قواعد البيانات النصية:

يعتبر النص الشق الأول في حمل المعلومات الرقمية على الويب، ويقابله شق المعلومات غير النصية الذي يتنوع بين معلومات سمعية ومرئية ثابتة أو متحركة. ويمكن القول إن الجزء الأكبر من قواعد البيانات البحثية المتاحة على الويب يوجد في الشكل النصي لعرض المعلومات، أما قواعد البيانات الوسائط الأخرى فقد خرجت لتخدم فئة أو مجالاً بحثيًا محددًا. مثل قواعد بيانات الصور الطبية والفلكية... وغيرها.

- ومن بين قواعد البيانات النصية: magportal [2].

٢ / ٢ / ٣ / ٢ قواعد بيانات الوسائط المتعددة:

يعتبر الباحث الويب الجزء المضيء من الإنترنت، وذلك لما تحمله الويب من وسائط مرئية مسموعة لحمل المعلومات وجعلها الأكثر جذبًا لمستخدمي الإنترنت.

---

(١) http://www.egypt.gov.eg

(٢) www.magprotal.com

ونتيجة لتنوع وسائط حمل المعلومات غير النصية، فمن غير الممكن أبدًا أن نجد قاعدة بيانات واحدة تملك ملفات لكل أشكال حمل المعلومات. ومن ثم فإن هناك قواعد بيانات الويب المتخصصة في المواد الصوتية، وقواعد البيانات المتخصصة في المواد المصورة الثابتة، وأخرى للمواد المصورة المتحركة، وغيرها للرسومات الحيوية... وقد انعكس التنوع والتباين أيضًا بين أشكال حمل المعلومات على عمل أدوات بحث الويب التي ذهبت إلى التخصص مثل محركات بحث الصور ومحركات بحث الصوت... إلخ.

ومثالاً لقواعد بيانات الصور: corbis [1]:

- قاعدة بيانات الفيديو: wolr of movies.
- قاعدة بيانات الصوت: All music [2].

٤/٢/٢ من حيث التخصص:

فرضت طبيعة الويب التخصص الموضوعي على قواعد البيانات، وذلك لما تحمله الويب من معلومات رقمية تخدم مختلف موضوعات المعرفة البشرية. وبلغ هذا الأمر الدرجة التي جعلت محركات البحث تتجه إلى التخصصات الدقيقة، على الرغم من أنها قواعد بيانات لمحتوى الويب عامة. وتبعًا لمحتوى قواعد بيانات الويب فإنها تنقسم إلى:

١/٤/٢/٢ قواعد بيانات متخصصة:

تتبع الكثير من قواعد البيانات هيئات علمية أو متخصصة في مجالات محددة، وعلى ذلك فإن قاعدة البيانات تحمل تخصص تلك الهيئة أو المؤسسة التابعة لها. ومن نماذج قواعد البيانات المتخصصة:

---

(١) www.corbis.com

(٢) http://www.allmusic.com

- قاعدة بيانات وكالة الفضاء ناسا: NASA [١].
- قاعدة بيانات الاتحاد الدولي لجمعيات للمكتبات والمعلومات: IFLA [٢].

٢/٤/٢/٢ قواعد بيانات عامة:

تعد محركات بحث الويب أفضل أمثلة قواعد البيانات العامة على الويب، بالإضافة لذلك توجد قواعد بيانات تعمل ككشافات للدوريات في مجالات مختلفة. ومن نماذج قواعد البيانات العامة:

- محرك بحث الويب العام Google [٣].
- قاعدة بيانات مقالات الدوريات find articles [٤].

٥/٢/٢ من حيث آليات البحث:

إن إحدى الوظائف الأساسية لقواعد البيانات هي استرجاع وبث النتائج إلى مستخدم قاعد البيانات بحسب حاجته الموضوعية. ولذلك توفر قاعدة البيانات آليات مختلفة لتمكين المستخدم من البحث والاسترجاع، وعلى الرغم من كثرة وتنوع آليات البحث التي تعتمد عليها قواعد البيانات، إلا أنها جميعًا ترتكز على أسلوبي البحث الحر بالكلمات المفتاحية (لغة الأوامر)، أو البحث بتصفح محتويات قاعدة البيانات.

١/٥/٢/٢ أسلوب البحث بالتصفح:

يستخدم أسلوب التصفح داخل قواعد بيانات الويب بأكثر من شكل ومسمى. فهناك شكل تصفح المصطلحات الدالة على موضوعات قاعدة البيانات ومن خلالها يتم استرجاع محتوى قاعدة البيانات، وفقًا للمصطلحات المقابلة لحاجة كل مستفيد، أيضًا هناك شكل التصفح بالوحدات المعلوماتية ذاتها كما في

---

(١) http://www.nasa.gov/multimedia/imagegallery/index.html

(٢) www.ifla.org

(٣) www.findarticles.com

(٤) www.findarticles.com

قواعد بيانات الصور عند استخدام لقطات الصور لعرض التقسيم الموضوعي لمجموعات الصور الرقمية. أما مسميات التصفح فتختلف بين البحث بالتقسيم الموضوعي وبين التصفح، ويرى الباحث أن الفارق بين المصطلحين يرجع في أغلب أحواله إلى اعتماد بعض قواعد البيانات لتقسيمات منهجية ترجع في منهجها إلى خطط تصنيف موضوعية أو مكانز. واعتمادًا على ذلك فإن استخدام قوائم رؤوس الموضوعات لتصفح محتويات قاعدة البيانات، إما أن يعتمد على لغة مضبوطة مقننة أو يعتمد على لغة طبيعية تعبر في المقام الأول عن محتوى قاعدة البيانات من زاوية واحتياجات مستخدمي الويب من زاوية أخرى.

إضافة إلى ذلك فهناك مجموعة من العوامل المساعدة التي يجب أن تؤخذ في الاعتبار عند الأخذ بأسلوب التصفح مثل سرعة تحميل الصور على شاشة عرض النظام ؛ حيث يتوقف على ذلك الوقت الذي يمكن فيه للمستفيد تصفح كامل قاعدة البيانات، سواء طال هذا الوقت أم قصر نتيجة لأداء نظام الاسترجاع. وكذلك حجم شاشة العرض وعدد الوحدات المسترجعة في كل شاشة عرض.

٢/٥/٢/٢ أسلوب البحث بالأوامر:

تستخدم قاعدة البيانات أسلوب الأوامر للتعامل مع مستخدمي الويب ذات المستوى المتقدم في مهارات استخدام قواعد البيانات. ولغة الأوامر هي لغة حوار مباشرة بين المستخدم والنظام، يوفر النظام من خلالها الاسترجاع السريع الدقيق المباشر لمحتويات قاعدة البيانات. وتعتمد لغة الأوامر في قواعد البيانات على آليات تسمى باستراتيجيات البحث ؛ حيث تستخدم هذه الآليات ومعاملاتها لرسم حدود واتجاهات موضوع البحث. ومن أشهر آليات البحث على الويب البحث البولياني Boolean searching الذي يعتمد على معاملات ثلاث أساسية هي and, or, not وآلية البحث بالحقول الواصفة field searching. وقد تستخدم قاعدة البحث فراغ البحث search box أو تستخدم قوائم منسدلة تساعد المستخدم أكثر في التمرس على لغة الأوامر.

ومن أبرز قواعد البيانات على الويب التي تستخدم لغة الأوامر، قاعدة بيانات ProQuest ؛ حيث تعتمد على أشكال مختلفة لاستخدام لغة الأوامر(١):

- استخدام الكلمات المفتاحية أو المصطلحات المضبوطة.
- استخدام البحث بالمنطق البولياني.
- البحث بحقول التسجيلات الببليوجرافية.
- البحث بالتقارب بين الكلمات، وهو يلائم أكثر لقواعد بيانات النصوص الكاملة.

٦/٢/٢ من حيث الإتاحة:

تحدد درجة الإتاحة لكل قاعدة بيانات بحسب مستوى الوصول الذي تسمح به قاعدة البيانات لمستخدم الويب. وتعتمد درجة إتاحة محتوى قاعدة البيانات على عوامل مختلفة، مثل:

أولاً: الطبيعة التجارية لموقع قاعدة البيانات.

ثانيًا: طبيعة المحتوى وحساسية الاطلاع عليه.

ثالثًا: الصفة الرقابية للمواقع الرسمية الحكومية.

ويمكن القول إن معيار الإتاحة هو العامل الأول الذي جعل من قواعد البيانات صعبة المنال لمحركات بحث الويب. حيث قد تفرض بعض قواعد البيانات تعليمات تمنع برنامج الزاحف من الوصول إلى ملفاتها، فضلاً عن نسخ تلك الملفات.

١/٦/٢/٢ قواعد بيانات محددة الوصول:

إن مصطلح محدودية الوصول أو الإتاحة ينطوي على أكثر من مستوى لحجب المحتوى عن مستخدمي الويب. وتتعدد مستويات الحجب بحسب طبيعة موقع

---

(١) Beck, Susan E. Searching web-based databases. Information literacy, 2007,cited at 30/6/2007, cited at http://lib.nmsu.edu/resources/dbpqd.html

قاعدة البيانات ؛ فمن قواعد البيانات ما تمنع المحتوى حتى دفع مقابل مادي للاشتراك بها subscription fee ، وأخرى حتى يتم تسجيل بيانات الاشتراك وبيانات المستخدم، وقاعدة ثالثة تمنع المحتوى إلا عن هوية مستخدم محدد. غير أن أغلب قواعد البيانات تطلب مقابلاً ماديًا للاطلاع على محتوياتها. ويرى Charles G. Doe أن هناك ملامح إيجابية تعطي الأفضلية لاستخدام قواعد البيانات ذات المقابل المادي. وتتلخص الإيجابيات في اتساع محتوى قاعدة البيانات وشمولية المحتوى، إلى جانب لجوء تلك القواعد لمستوى التحكيم العلمي الذي يضمن معلومات دقيقة، كما أن قواعد البيانات هذه تساعد مستخدمي الويب على التمرس ببث محتوى قواعد البيانات[1].

٢/٦/٢/٢ قواعد بيانات مفتوحة:

تتسم قواعد البيانات المفتوحة بتقديم محتواها من ملفات المعلومات دون مقابل مادي بما يتوافق والغالب من مستخدمي الويب. غير أن هذه المواقع قد تسبب تشتتًا كبيرًا في استرجاع المعلومات ؛ حيث تعمل هذه القواعد وكأنها مواقع إعلانية تركز أكثر ما تركز على استدعاء كم من الإعلانات يسبب بلبلة في استرجاع المعلومات. فضلاً عن أن هذه القواعد قد تحمل محتوىً غير محكم أو دقيق بالقدر الكافي في تناول التخصص.

٣/٢ بناء قواعد الويب وتصميمها :

تهدف قواعد البيانات إلى اختزان البيانات ومعالجتها، ثم تقديمها مرتبة وفقًا لنسق معين بما يتوافق واحتياجات مستخدميها. ولهذا تتعدد أشكال تصميم وبناء قواعد البيانات للوصول إلى أفضل أداء لمعالجة البيانات. وبصفة عامة فإن قواعد البيانات تعتمد في بنيتها على نظام تشغيل وإدارة قاعدة البيانات DBMS مثل SQL, DB2، واعتمادًا على قوة وبنية النظام المتبع تكون قدرة ومستوى أداء

(١) Doe, Charles G. Web-based databases and search tools, information today, Inc, 2007, cited at 30/6/2007, cited at http://www.infotoday.com/MMSchools/sep04/doe.shtml

قاعدة البيانات في معالجة وتنظيم وبحث البيانات. وتوجد ثلاثة أشكال أساسية تبنى عليها قاعدة البيانات؛ فإما أن تكون قاعدة البيانات ذات بنية هرمية أو شبكية أو علائقية. والاختيار بين هذه الأشكال يعتمد بالدرجة الأولى على طبيعة المؤسسة والتخصص الذي تهدف قاعدة البيانات لمعالجة البيانات فيه ؛ فيمكن لقواعد البيانات الهرمية أن تكون ذات الفائدة الأكبر من الأشكال الأخرى في حالة التعامل مع بيانات مؤسسة تتكون من أكثر من مستوى إداري وكل مستوى له فروعه المختلفة. أما شكل قواعد البيانات العلائقية فيكون الأكثر إفادة في حالة التعامل مع بيانات لها روابط متداخلة فيما بينها كبيانات العملاء والسلع والمعاملات لمؤسسة تجارية. والحديث عن تصميم قواعد البيانات هو ذاته الحديث عن خصائص نظم إدارة قواعد البيانات، لأن قاعدة البيانات ذاتها إنما تتكون من شقين ؛ أولهما : البيانات، وثانيهما : تعليمات المعالجة والتشغيل. ويختلف الشق الأول بحسب المجال فهناك البيانات الببليوجرافية للمكتبات والمعلومات، وهناك النصوص الكاملة ... وغيرها. أما الشق الثاني فيضطلع بأدائه نظام إدارة قاعدة البيانات بحسب تعليمات التنظيم داخل الجداول والبحث والاسترجاع.

١/٣/٢ التصميم العام لقواعد البيانات:

ينطوي التصميم العام لقواعد البيانات على كونها وعاءً يحوي مجموعات من البيانات التي تحمل فيما بينها خصائص متشابهة. وهي بذلك تختلف عن ورقة البيانات spreadsheet التي تهدف فقط إلى أداء مجموعة من العمليات الحسابية على مجموعة صغيرة من البيانات. ومن الخطأ هنا محاولة التشبيه بين كل من قاعدة وورقة البيانات، فهما يستخدمان الأعمدة والصفوف في تنظيم البيانات غير أن هناك اختلافات جوهرية تعكس مدى التفاوت بين وظيفة كل منهما. ويوضح الجدول (٢-١) جوانب الاختلاف من خلال مقارنة قاعدة وورقة البيانات[١].

(١) Marc Grange. Databases design, the author, 2006, cited 5/7/2007, cited at http://www.marc-grange.net/BDConception.htm

الجدول (٢-١) أوجه الاختلاف بين بنية قاعدة وورقة البيانات

أوجه الاختلاف	قواعد البيانات	ورقة البيانات
الوظيفة الرئيسية	معالجة وتنظيم البيانات	العمليات الحسابية
بنية البيانات	بنية وتخزين ثابتين	لا توجد
تكامل ووحدة البيانات	التعرف الدقيق على قيمة البيانات	لا توجد
إتاحة البيانات	مستخدمون متعددون	مستخدم واحد
السرية والخصوصية	مراجعة حق الوصول لكل مستخدم	لا يوجد تحكم
حجم البيانات	تتكون من جداول متعددة يمكنها احتواء الآلاف من صفوف البيانات	جدول واحد من مجموعة صفوف
المعالجة والتنظيم للبيانات	تقوم بالمعالجة الكمية والنوعية	معالجة كمية فقط
استرجاع البيانات	تستخدم تعليمات البحث SQL	تستخدم فقط لأشخاص محددين

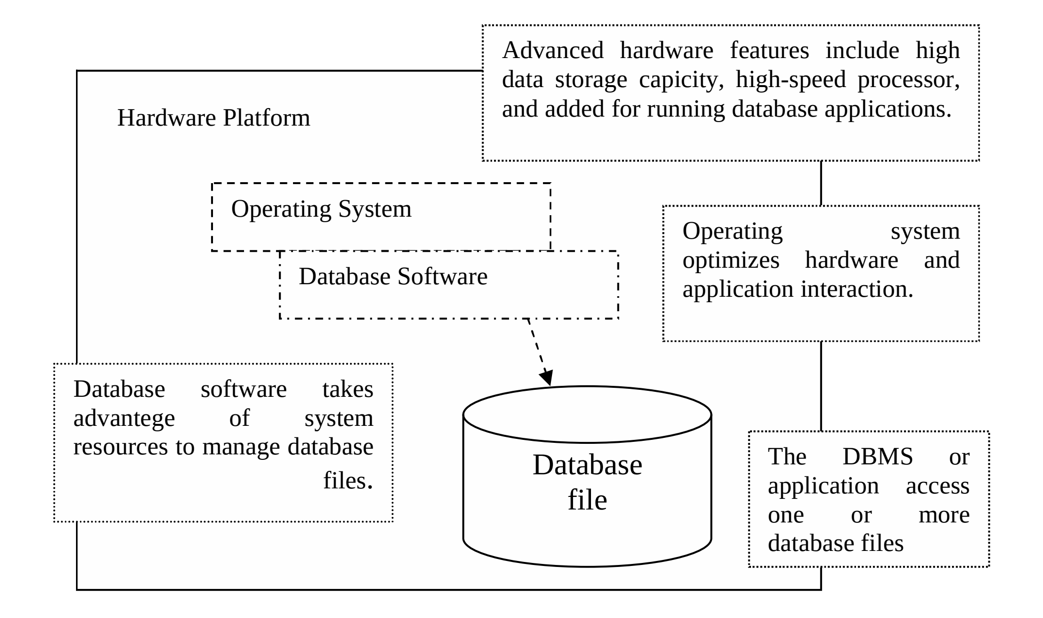
يوضح الجدول (٢-١) بعض خصائص العمل بقواعد البيانات كما يلي:

- أن الهدف الرئيس لعمل قاعدة البيانات هو تنظيم ومعالجة مجموعات البيانات حتى يمكن إخراج معلومات ذات معنى محدد يتم استرجاعها بحسب الحاجة.
- أن قاعدة البيانات ليست تجميعًا فقط للبيانات، وإنما تقوم بعمليات منطقية وكمية على هذه البيانات لا تستطيع أوعية التخزين الأخرى القيام بها.
- تعمل قاعدة البيانات على تنظيم إجراءات الإتاحة وحماية الوصول إلى البيانات المختزنة داخلها.

يتكون التصميم العام لقاعدة البيانات من مجموعة من العناصر التي تساهم جميعًا في إتمام مهام قاعدة البيانات. وحالها في ذلك حال مختلف نظم الاسترجاع ؛ حيث تعتمد قاعدة البيانات على ما يلي :

أولاً : المكونات المادية والأجهزة hard ware : وهي الخادم ذو الإمكانات العالية في التخزين والتوافق بالكفاءة مع نظم التشغيل المستخدمة.

ثانيًا: البرامج والتطبيقات التي تعمل بها قاعدة البيانات وتتمثل في عنصر واحد هو نظم إدارة قواعد البيانات DBMS. ويزيد على هذه المكونات عنصر برامج ولغات التوافق مع الويب Sceipting language ، في حالة تناول قواعد البيانات على الويب. وقد بلغت نظم إدارة قواعد البيانات الدرجة التي ينبغي لمن يتناول تصميم قواعد البيانات أن يتحدث فقط عن تصميم وعمل نظم إدارة قواعد البيانات. ويوضح الشكل (٢-٥) التصميم والمكونات العامة لتشغيل قاعدة البيانات(١):



الشكل (٢-٥) التصميم والمكونات العامة لتشغيل قاعدة البيانات

يوضح الشكل (٢-٥) أن قاعدة البيانات تعتمد على مجموعة من العناصر المتكاملة يساهم كل منها في أداء دور محدد ، وهذه العناصر تفصيلاً هي :

(١) Dave Hillman. Relational databases , the author, 2005,cited 17/2007, cite at http://dhillman.com/theplace/databoot/databases.htm

أولاً: المكونات المادية:

١- أقراص صلبة لتخزين كميات كبيرة من البيانات.

٢- معالجات ذات قدرة عالية، وسرعة تلائم تشغيل قاعدة البيانات.

٣- أجهزة أخرى بحسب تطبيقات قاعدة البيانات؛ كأجهزة تشغيل الملفات المسموعة أو المرئية.

ثانيًا: برامج إدارة قاعدة البيانات:

١- نظم التشغيل Operating systems.

٢- نظم إدارة قاعدة البيانات DBMS.

وتختلف هذه البرامج بحسب طبيعة وبيئة تشغيل قاعدة البيانات، حيث تنقسم هذه النظم إلى:

أ-نظم إدارة قواعد بيانات سطح المكتب Desktop platforms .

ب- نظم إدارة قواعد بيانات بيئة الشبكات أو الويب Network servers.

ج- نظم إدارة قواعد بيانات الخادمات Database servers.

د- برامج ذات تطبيقات مختلفة وفقًا لمهام قاعدة البيانات.

بعد أن يستقر فريق عمل قاعدة البيانات على المكونات السابقة، يبدأ العمل لتنفيذ تصميم قاعدة البيانات وفقًا للخطوات التالية[١]:

- **الخطوة الأولى** : تحويل مختلف البيانات المراد معالجتها إلى عناصر ومداخل.

- **الخطوة الثانية** : الوقوف على الروابط والعلاقات المنطقية بين عناصر البيانات حتى يتم تفعيل هذه الروابط في شكل علاقات بين حقول القاعدة وجداولها. ERD Entity-Relationship Diagrams

(١) System analysis and design ; chapter 7. http://aslnx1.uah.edu/~floydj/MIS497/chapter07.ppt

- **الخطوة الثالثة:** إنشاء عناصر تطبيع قاعدة البيانات Database Normalization؛ من خلال الاهتمام بتجنب التكرار والتداخل بين البيانات.
- **الخطوة الرابعة** : التعريف بدليل مختلف مداخل البيانات والحقول.
- **الخطوة الخامسة** : بعد الانتهاء من تصميم الروابط المنطقية بين البيانات وإعداد الجداول وتطبيع القاعدة، تحول كل هذه العناصر إلى قاعدة بيانات.

ويضيف الباحث إلى الخطوات السابقة خطوات أخرى تخص قواعد بيانات الويب، مثل :

- **الخطوة السادسة:** تصميم واجهات موقع قاعدة البيانات databases interface.
- **الخطوة السابعة** : رسم آليات التواصل بين مستخدم الويب وقاعدة البيانات database connectivity : حيث الاختيار بين آليتي (Two-tier OR Three-tier Architecture)

ينتهي تصميم قواعد بيانات الويب على اختلاف محتواها وفقًا للخطوات السابقة. غير أن درجة الاهتمام ببناء قاعدة بيانات الويب تؤثر بدورها على طبيعة أداء قاعدة البيانات ؛ حيث تتسم قواعد بيانات الويب بملامح عمل يجب أن تسير وفقًا لما يرجوه مستخدمو الويب، مثل(١) :

❖ الإتاحة العالمية :Global access

تستطيع الويب تقديم الإتاحة العالمية أو الكونية لقاعدة البيانات؛ لما تمتلكه من تقنيات متطورة في الاتصال عن بعد، إلى جانب التمتع بمعايير نقل البيانات خلال مدى مكاني كبير ومدى زمني قصير. ولا يتطلب الوصول إلى قاعدة البيانات سوى متصفح للويب وخط اتصال بالإنترنت.

---

(١) What are the characteristics of a web-based database design? Blurtit.inc, 2007, cited at 2/7/2007, cited at http://www.blurtit.com/ask.php

❖ سهولة الاستخدام :Ease of use

يقدم متصفح الويب قاعدة البيانات إلى مستخدمي الويب وفقًا لطبيعة التعامل المتآلف معها في مواقع الويب.

❖ ملاءمة التقنيات المتعددة : Multiple platforms

تستطيع تقنيات الويب التعامل مع مختلف أنواع الأجهزة المادية والبرامج الداخلة في تصميم وبناء قواعد البيانات؛ مما يوفر كفاءة أداء قواعد بيانات الويب على اختلاف نظم إداراتها.

❖ فاعلية التكلفة :Cost effectiveness

حيث تحتاج قواعد بيانات الويب فقط إلى متصفح ويب وخط اتصال للإنترنت، كي يستطيع مستخدم الويب الحصول على حاجاته المعلوماتية.

❖ جوانب آمن المعلومات :Security issues

تختلف درجة حساسية المعلومات المقدمة داخل قواعد البيانات، التي تصل إلى مداها في قواعد بيانات الهيئات الحكومية والأمنية. ومن ثم كان لزامًا الاهتمام بحماية البيانات وسهولة الوصول في الوقت ذاته لمن لهم حق الإتاحة.

❖ جوانب التوافق : Adaptability issues

تقدم الويب ميزات متعددة لمستخدمي الويب نتيجة لما تحققه من توافق مع الأنظمة المختلفة لقاعدة البيانات ؛ فهي تحول الواجهات التقليدية لقواعد البيانات إلى واجهات الويب الأكثر اعتيادًا واستخدامًا، على الرغم مما قد يتطلبه ذلك في توفير أجهزة وتقنيات مختلفة لتقديم قواعد بيانات الويب على الشكل الذي يحقق كفاءة أداء عالية.

١/١/٣/٢ تطبيع قاعدة البيانات : Database Normalization

أولاً: مفهوم التطبيع :

ينطوي مصطلح تطبيع قاعدة البيانات Database Normalization على مفهوم مؤداه أن تؤدي قاعدة البيانات مهامها بالشكل الطبيعي دون أي خطأ ؛ مثل أخطاء تداخل البيانات أو تكرارها. ومن ثم فإن لعملية تطبيع قاعدة البيانات هدفين أساسيين هما:

١- تحديد عناصر البيانات eliminating ذات الكم الكبير بما يمنع معه تسجيل بيانات مترابطة أو متجانسة في جداول مختلفة، ومن ثم تداخل البيانات.

٢- التحكم في مجموعات البيانات، حيث تكون كل مجموعة مترابطة من البيانات مسجلة في جدول واحد.

ثانيًا: أشكال تطبيع قاعدة البيانات: Normal Forms

عند تحقيق هذين الهدفين فإن قاعدة البيانات يمكنها أداء مختلف مهامها دون التعرض لعمليات أو نتائج خاطئة. وتمر عملية تطبيع قاعدة البيانات بأشكال أو مراحل مختلفة تحدد في خمسة أشكال أساسية تبدأ من **1NF** حتى **5NF**. غير أن الشكل الخامس قلما يستخدم في تصميم قواعد البيانات. وتعمل الأشكال الأربعة الأساسية كملامح أو قواعد إرشادية لتحسين تصميم وبنا قواعد البيانات وهذه الأشكال كالتالي[١]:

الشكل الأول (Normal Form 1):

يحتوي الشكل الأول على القواعد الأساسية لبناء قاعدة البيانات في:

(١) Chapple, Mike. Database normalization basics, about.inc, 2006, cited at 7/7/2007, cited at http://databases.about.com/mbiopage.htm

١- يجب تحديد عناصر البيانات بما يلغي معه تكرار عمودين لعنصر واحد داخل جدول البيانات.

٢- إنشاء جدول منفصل لكل مجموعة مترابطة من البيانات داخل قاعدة البيانات.

الشكل الثاني (Normal Form 2):

١- يعتمد الشكل الثاني على تنفيذ القواعد الواردة في الشكل الأول، حيث يتم :

٢- حذف مجموعات البيانات الفرعية المتداخلة والمكررة داخل الجدول الواحد، ووضعها في جداول مختلفة.

٣- تفعيل الروابط المنطقية بين مجموعات البيانات، وإنشاء العلاقات بين الجداول المترابطة، وعمل المفاتيح الأساسية.

الشكل الثالث (Normal Form 3):

١- يبنى الشكل الثالث على التفعيل الصحيح لقواعد الشكل الثاني.

٢- يؤكد الشكل الثالث على انتفاء كل الأعمدة والعناصر التي لا تعتمد على مفاتيح أساسية أو مرتبطة بحقول مفاتيح أساسية. وهي تلك الحقول التي لا ترتبط بأي من مجموعات البيانات.

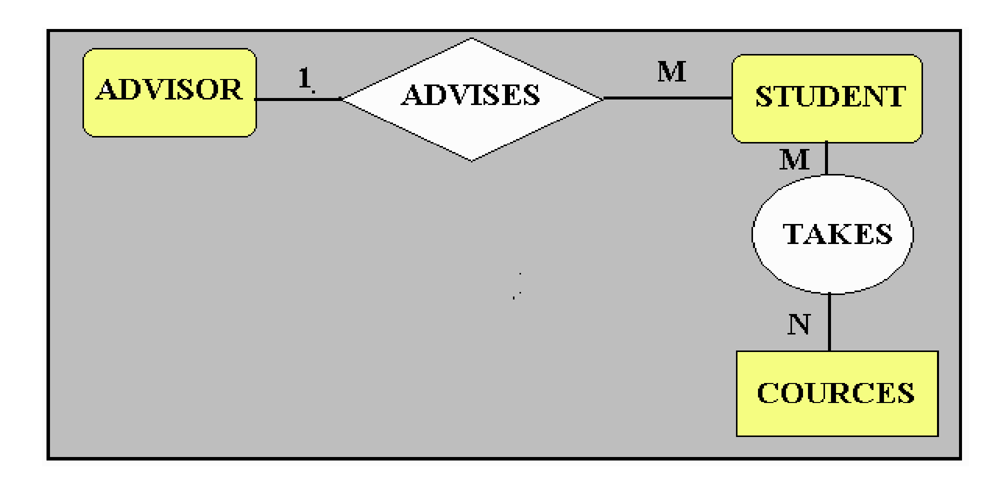
الشكل الرابع (Normal Form 4):

١- يبنى الشكل الرابع على التفعيل الصحيح لقواعد الشكل الثالث.

٢- يقوم الشكل الرابع بتفعيل العلاقات بين الجداول حتى تتمكن قاعدة البيانات من تنظيم معالجة البيانات، ثم استرجاعها بشكل دقيق.

مما سبق يمكن القول إن عملية تطبيع قاعدة البيانات هي جعل قاعدة البيانات تسير وفق المنهج الطبيعي لبناء قواعد البيانات. ويوضح الشكل (٢-٦)

نموذجًا لتطبيع قاعدة بيانات الإشراف التعليمي المكونة من مجموعات بيانات تتعلق بالمشرف والمقررات والطلاب.



الشكل (٢-٦) نموذج لتطبيع قاعدة بيانات

٢/٣/٢ نظم إدارة قواعد البيانات DBMS:

تقف نظم إدارة قواعد البيانات خلف الوظائف والمهام التي تؤديها قواعد البيانات ؛ حيث تمثل هذه النظم المحرك الفعلي لمختلف عمليات المعالجة والتجهيز التي تتم على البيانات والوحدات الداخلية. ويمكن الحديث عن هذه النظم من ناحيتي الأنواع والمكونات، كما يلي :

١/٢/٣/٢ أنواع نظم إدارة قواعد البيانات:

تعكس خصائص وملامح نظام إدارة قاعدة البيانات مدى الفاعلية والكفاءة التي تقدمها هذه القاعدة. وإذا أردنا الحديث عن تصميم وبناء قواعد البيانات والعوامل المؤثرة فيها ، فيمكن أن نتحدث عن نظم إدارة قواعد البيانات. ويعكس ذلك واقع بناء قواعد البيانات ؛ حيث إن قاعدة البيانات إنما هي نظام الإدارة، فسلبياتها سلبياته، وإيجابيتها إيجابياته. وتؤثر الأنواع والفئات المختلفة

من نظم إدارة قواعد البيانات في تقسيم فئات قواعد البيانات كما سبق ذكره في أنواع قواعد البيانات. وتقوم نظم إدارة قواعد البيانات بتنفيذ مختلف المهام الرئيسية لقاعدة البيانات، مثل :

- تقديم الحلول والعمليات الخاصة بمعالجة وتنظيم ملفات المعلومات.
- تنظيم وظائف القاعدة والتحكم في سير العمليات داخل قاعدة البيانات.
- توفير الإتاحة بحدودها المختلفة، وتوفير الوقت، والمرونة مع استفسارات وبحث مستخدمي القاعدة.

تتنوع نظم إدارة قواعد البيانات وفقًا للاحتياجات المختلفة في إنشاء قواعد البيانات ؛ فتوجد نظم إدارة قواعد بيانات سطح المكتب، ونظم قواعد بيانات الخادم، ونظم قواعد الويب. كما تختلف هذه النظم أيضًا فيما بينها من حيث شكل تصميم قاعدة البيانات، حيث توجد نظم قواعد البيانات العلائقية، ونظم قواعد البيانات الهرمية، ونظم قواعد البيانات الشبكية. ومن أكثر النظم استخدامًا لتصميم قواعد البيانات ما يلي:

الجدول (٢- ٢) أكثر النظم استخدامًا لبناء قواعد البيانات

نظم قواعد البيانات الشخصية	نظم قواعد البيانات المؤسسية
☒ **MS Access**	☒ **Oracle**
☒ **MS FoxPro**	☒ **MySql**
☒ **Filemaker Pro (Mac & Win)**	☒ **Informix**
☒ **Paradox**	☒ **IBM DB2**
☒ **Approach**	☒ **Microsoft Sql Server**

يوضح الجدول (٢-٢) أن اختيار أي من النظم السابقة يعتمد على طبيعة حجم قاعدة البيانات المراد بناؤها. إضافة إلى تحديد حجم الأداء المتوقع في معالجة كم

محدد من المعلومات. وفيما يلي تناول الخصائص الإيجابية والسلبية لأهم نظم إدارة قواعد البيانات التي وردت بالجدول (٢-٢)[١].

أولاً: قاعدة البيانات MS Access:

يستخدم برنامج MS Access لتحقيق المهام المحدودة والبسيطة في عمل قواعد البيانات صغيرة الحجم، كما يستخدم كقاعدة بيانات في المواقع غير التجارية أو مواقع الهواة، أو تلك المواقع التي لا تتوقع معالجة وتشغيل كميات كبيرة من مجموعات المعلومات.

ميزات MS Access:

١- قاعدة بيانات علائقية تحقق قدرًا كبيرًا من الكفاءة والسهولة في معالجة البيانات.

٢- تدعم استخدام البيانات المصورة أو الأشكال الرسومية.

٣- قليلة التكلفة أو متاحة للتحميل.

٤- سهلة التعلم والتعامل.

٥- سريعة الأداء في معالجة البيانات.

سلبيات MS Access:

١- لا تقبل التعديل أو الإضافة دون وجود ترخيص البرمجة Software licences.

٢- قابلة للسقوط أو للأداء الخاطئ إذا لم تستخدم جيدًا على خادم الويب.

٣- تستخدم في بيئة الويندوز فقط.

٤- ذات قدرات محدودة بالمقارنة مع القواعد الأخرى التي تستخدم لغة SQL.

---

(١) Databases. Find my hosting, 2002, cited at 10/7/2007, cited at http://www.findmyhosting.com/index.asp

٥- يضعف أداؤها مع مجموعات البيانات ذات الحجم الكبير.

٦- أقل ثقة في قدراتها من قواعد البيانات الأخرى.

ثانيًا: قاعدة البيانات MY SQL:

تستطيع قاعدة البيانات MY SQL التوافق مع نماذج مختلفة من أنظمة التشغيل وبيئات عمل قواعد البيانات. كما تتمتع هذه القاعدة بالأداء القوي المتوافق مع أنظمة عمل مختلفة. وتشترك هذه القاعدة مع لغة التحويل PHP لإخراج واجهة موقع الويب الممثل لقاعدة البيانات.

ميزات MY SQL:

١- يمكن الحصول عليها بسهولة **open source**، نظرًا لإتاحة برامجها.

٢- قاعدة بيانات ذات كفاءة عالية واستقرار في الأداء بما يحقق الثقة في النتائج المسترجعة.

٣- تقوم بمعالجة البيانات بمختلف أنواعها، كما تستطيع التوافق مع الأحجام المختلفة من البيانات.

٤- تقدم معالجة سريعة لكميات البيانات.

٥- تستطيع معالجة كميات كبيرة من مجموعات المعلومات.

سلبيات MY SQL:

١- صعبة الاستخدام لحديثي التعامل مع قواعد البيانات.

٢- هناك جدل حول مدى ارتباطها بقواعد محددة.

٣- لا تستطيع استخدام الأنماط المعقدة من الروابط المعقدة ذات التداخل متعدد المستوى.

٤- لا تستطيع اختزان الإجراءات بما يمكنها من الرجوع لخطوات أو إجراءات سابقة.

ثالثًا: قاعدة البيانات SQL Server:

تعتبر هذه القاعدة أحد الأنواع المهمة في التعامل مع كميات البيانات كبيرة الحجم. ويمكن القول إنها اللغة المقيدة للعمل في مختلف أنواع قواعد البيانات العلائقية ؛ حيث تؤدي هذه القاعدة مختلف المهام الرئيسة في التعامل مع البيانات من إنشاء وإضافة وتحديث وصياغة لأسئلة البحث.

ميزات SQL Server:

١-ذات كفاءة عالية في معالجة كميات كبيرة من البيانات.

٢-تدعم إجراءات العمل المعقدة، بالإضافة إلى دعم آليات تحليل البيانات.

سلبيات SQL Server:

١- عالية التكلفة.

٢-تحتاج إلى دراية بالغة بقواعد عملها.

رابعًا: قاعدة البيانات Oracle:

تملك قاعدة البيانات Oracle شهرة كبيرة بين قواعد البيانات ؛ وذلك لما تتصف به من قدرة على التوافق مع أنظمة تشغيل مختلفة، كما أنها تعالج مجموعات كبيرة من البيانات، فضلاً عن أداء مختلف مهام قواعد البيانات بقدرة عالية.

ميزات Oracle:

١-تملك أنواعًا من البيانات تستطيع إدخالها مباشرة إلى النظام.

٢-تملك قدرات التعامل مع كميات هائلة من مجموعات البيانات.

٣-تملك قدرة عالية في التوافق مع خادم قاعدة البيانات.

٤-أحد قواعد البيانات العلائقية التي تسهل معالجة البيانات.

سلبيات Oracle:

١- عالية التكلفة بالقدر الذي جعلها نادرة في مواقع الإنترنت ذات المستوى المتواضع.

٢- تحتاج إلى مدير قاعدة بيانات عالي المهارة، ذي مستوى متقدم في آلية عمل قواعد البيانات.

يعتمد الاختيار بين نظم بناء قواعد البيانات السابقة على عددٍ من العوامل مثل؛ الاستخدام قصير أم طويل الأجل لقاعدة البيانات، أيضا حجم مجموعات البيانات التي سيتم معالجتها داخل قاعدة البيانات، وطبيعة موقع الويب وواجهة أداء قاعدة البيانات، وحجم الاستخدام اليومي عدد الإجراءات التي تقوم بتنفيذها قاعدة البيانات مع كل مستخدم، إلى جانب مستوى مدير قاعدة البيانات الذي يختلف بين مستويات المهارة المتقدمة وحديثي التعامل مع قواعد البيانات، ويضاف إلى ذلك جميعًا عامل التكلفة الذي يبدأ بشراء برنامج تصميم قاعدة البيانات ثم تأجير أو شراء مساحة تخزينية داخل خادم قاعدة البيانات database server.

٢/٢/٣/٢ مكونات نظم إدارة قواعد البيانات:

تتكون نظم إدارة قواعد البيانات على اختلافها من عناصر فرعية، تتكامل معًا لتحقيق مهام ووظائف قاعدة البيانات في معالجة البيانات وتنظيمها واسترجاعها. وهذه المكونات تتمثل في:

المكون الأول: واجهات قاعدة البيانات، مدير قاعدة البيانات، نظم معلومات مرتبطة.

يعتمد المكون الأول في بناء نظام إدارة قاعدة البيانات على ثلاث ركائز هي:

- **واجهات قاعدة البيانات Users Interfaces :**
  - لغة البحث بالأوامر.
  - لغة البحث بالتصفح أو الأمثلة.
  - لغة الاستفسارات SQL.

- **مدير قاعدة البيانات Database Administrators** : ويتولى مهام إدارة الوظائف، وصيانة قاعدة البيانات.

- **نظم معلومات مرتبطة بقاعدة البيانات:**

تدعم نظم إدارة قواعد البيانات بعض البرامج الأخرى الملائمة لإدخال وإخراج مجموعات البيانات.

المكون الثاني : لغة معاملة البيانات Data Manipulation Language:

تقوم لغة معاملة البيانات **DML** بأداء وظائف الإدخال والتعديل والتحديث والإضافة والحذف لمجموعات البيانات داخل قاعدة البيانات.

المكون الثالث : مخطط أو شكل قاعدة البيانات Database Schema:

يطلق مصطلح مخطط أو شكل قاعدة البيانات على التصميم الكامل لقاعدة البيانات المكون من الحقول والتسجيلات والملفات بالإضافة إلى الروابط بين الجداول داخل قاعدة البيانات، ويمكن أن يطلق على أي منها مخططات فرعية **subschemas**.

المكون الرابع : مستودع تخزين البيانات:

- تتجه مجموعات البيانات على اختلاف جداولها وتوزيعها إلى مستودع اختزان البيانات.
- قد يكون هذا المستودع مركزًا أو موزعًا على مناطق متفرقة كما في حالة بعض قواعد البيانات.

تتفق نظم إدارة قواعد البيانات على اختلاف أنواعها في المكونات السابق ذكرها، غير أن ثمة جوانب يجب مراعاتها عند الاختيار بين نظم إدارة قواعد البيانات، وتتمثل في[1]:

(١) Web Databases. MDA, 2002,cited at 20/6/2007, cited at http://www.mda.org.uk/webdatabases.htm

١- التنصيب أو التركيب Installation:

يجب النظر إلى جوانب المرونة والتعقيد في تنصيب النظام على خادم الويب، بالإضافة إلى مدى التوافق أو الاختلاف بين نظام عمل الخادم ونظام إدارة قاعدة البيانات.

٢- الصيانة Maintenance:

ينبغي الاهتمام بصيانة نظام قاعدة البيانات أثناء عملها يومًا بعد آخر، إلى جانب السؤال عمن تقع مسئولية صيانة نظام عمل قاعدة البيانات عليه؛ هل على الشركة المنتجة أم على هيئة صيانة أخرى.

٣- عامل التكلفة cost:

تتفاوت نظم إدارة قواعد البيانات في تكاليفها بين المتاح في استخدامه وبين ما يتكلف الآلاف لتجديد اشتراكه سنويًا، لذلك ينبغي اختيار افضل النظم تأديةً للمهام قاعدة البيانات وتوفيراً للنفقات.

٤- ترخيص الاستخدام Licensing:

تحدد بعض تراخيص استخدام نظم قواعد البيانات بعض الشروط القاسية التي قد تحد من الاستخدام الأمثل لقاعدة البيانات، خاصة في حالة إتاحتها على الويب.

٥- مكان التنصيب والتركيب Hosting:

تواجه بعض قواعد البيانات مشكلة رفض خادم الويب لضم قاعدة البيانات لموقع الويب التابعة له. أما في حالة القبول فإنه يجب تحقيق التوافق بين نظام عمل خادم الويب ونظام إدارة قاعدة البيانات.

٦- أمن البيانات Security:

تتطلب حماية البيانات إرساء حدود لمستويات الوصول والتعامل مع برامج قاعدة البيانات، فضلاً عن تلك البرامج التي تخص خادم الويب المضيف لقاعدة البيانات.

٧- النقل والتحويل Migration:

تتحول بعض قواعد البيانات من حالة الاستخدام المحلي إلى الخط المباشر، وهذا يتطلب اختيار أحد نظم إدارة قواعد البيانات على الخط المباشر التي يمكنها استيراد كل بيانات القاعدة المتوقفة بسهولة ويسر. ومن أهم الأنظمة التي تؤدي ذلك نظام MySQL.

وجدير بالذكر هنا هو قدرة نظام إدارة قاعدة البيانات على الويب في بيئة الويب، إلى جانب التوافق بين نظام العمل ولغات التحويل حتى تؤدي قاعدة البيانات وظائفها وفق ما اعتاد عليه مستخدمو الويب. ولا يجب أن يشغلنا تناول البيانات النصية بما ينسينا معه المصادر غير النصية على الويب. ولما تملكه قواعد بيانات الوسائط المتعددة على الويب، فإن نظم إدارة قواعد بيانات الوسائط المتعددة على الويب MMDBMS تحتاج إلى توافر عدد من الخصائص، حتى تستطيع القيام بمعالجة وتنظيم المواد السمعية البصرية كما في حالة المواد النصية. وتعمل هذه النظم مع قواعد البيانات التي تملك تقنيات مثل Active Server Pages, ColdFusion and Java Server Pages. ومن هذه الخصائص ما يلي :

❖ **خصائص التشغيل :**

أن تستطيع هذه النظم أداء مختلف وظائف قواعد البيانات من إضافة وحذف وتعديل وتحديث، على مختلف أنواع البيانات غير النصية كالصوت والصورة الثابتة أو المتحركة.

❖ **مطابقة المعايير :**

يجب أن تؤدي هذه النظم وظائفها بمعايير الإجادة نفسها لنظم إدارة قواعد البيانات غير النصية، مثل مهام التنظيم والمعالجة والبحث والتحكم في الإجراءات.

❖ **قدرات العمل :**

تحتاج هذه النظم إلى التعامل مع كميات كبيرة من مجموعات البيانات غير النصية، مهما كانت أحجام تلك المجموعات أو مساحتها التخزينية.

❖ **مرونة الأداء :**

يجب أن تسمح هذه النظم لمستخدمي قاعدة البيانات بأداء مختلف العمليات مثل البحث وتجميع البيانات.

❖ **تقنيات أداء متقدمة :**

تعتمد نظم إدارة قواعد البيانات غير النصية على استخدام تقنية الاسترجاع المبني على المحتوى CBDR، بجانب امتلاكها لخصائص الاسترجاع النصي TBDR. وذلك وفقاً لنوع البيانات الذي تعالجه.

❖ **التوافق مع البيانات :**

تحتاج هذه النظم إلى ملامح خاصة في عمليات المعالجة والتكشيف والاسترجاع والتحليل، لما تملكه من أنواع بيانات مختلفة تتطلب عمليات وشاشات عرض متنوعة.

❖ **ملامح متقدمة :**

يجب أن تعمل هذه النظم بملامح متقدمة ترتقي إلى إمكانية إدراك الصوت والصورة وتحليل البيانات المصورة.

❖ **الوسائط الفائقة :**

تحتاج هذه النظم إلى فهم طبيعة القفز الفائق hypermedia databases والتنقل بين عقد nodes تجمع البيانات، خاصة عند العمل مع قواعد البيانات الموزعة.

٣/٢/٣/٢ أشكال نظم إدارة قواعد البيانات :

تملك نظم إدارة قواعد البيانات هياكل بناء مختلفة، تؤثر بدورها على تنظيم البيانات ومعالجتها داخل جسم قاعدة البيانات. ويمكن القول إن شكل نظام

قاعدة البيانات هو تخطيط وهيكلة تدفق البيانات data schema وتنظيمها داخل مستودعات قواعد البيانات. وتقسم قواعد البيانات هنا بحسب تخطيط قاعدة البيانات والشكل الذي ستؤدي مهامها بناءً عليه، غير أن اختيار أحد أشكال قواعد البيانات هنا لا يعتمد على كم البيانات المعالجة أو حجم استخدامها، وإنما يعتمد على طبيعة المؤسسة أو النشاط الذي تعمل لأجله قاعدة البيانات. وتقسم الأشكال المتاحة للاختيار إلى ثلاثة أشكال أساسية.

أولاً: شكل قواعد البيانات العلائقية.

ثانيًا: وقواعد البيانات الهرمية.

ثالثًا: قواعد البيانات الشبكية.

ولكل من هذه الأشكال قدراتها وخصائصها التي تتحكم في اختيارها أو تركها بحسب احتياجات معالجة البيانات. ويمكن شرح هذه الأشكال كما يلي[1]:

الشكل الأول : قواعد البيانات العلائقية RDBMS:

تعرف قواعد البيانات العلائقية Relational Database Management System على أنها قواعد بيانات تم بناؤها باستخدام آليات الربط (العلاقات) بين جداول قاعدة البيانات. وترتكز قواعد البيانات العلائقية على مفاهيم بنية البيانات والاختزان والاسترجاع إلى جانب تكامل مجموعات البيانات. وفي هذا الشكل تنظم مجموعات البيانات والعلاقات فيما بينها في شكل جداول مكونة من مجموعة من التسجيلات، وتتكون كل تسجيلة داخل الجدول الواحد من مجموعة الحقول نفسها المكونة لكل تسجيلات الجدول. وترمز مختلف حقول التسجيلات بعلامات مخصصة تسمى المفاتيح الأساسية Primary keys، وتعمل هذه المفاتيح على دعم التكشيف والبحث لقيم حقول التسجيلات. وتستخدم آلية الربط بالعلاقات بين الحقول، عندما يوجد حقلان أو أكثر في جداول مختلفة،

(١) Practical database design, IBM, 2003, cited at 25/6/2007, cited at http://www.ibm.com/developerworks/web/library/wa-dbdsgn1.html

غير أنهما يحملان قيم البيانات نفسها، وتجرى عملية الربط بينهما حتى يمكن مضاهاة تسجيلات البيانات وتحديد مدى الترابط بين التسجيلات في الجداول المختلفة. وغالبًا ما تحمل الحقول ذات العلاقات أسماءً متشابهة. ومثال على ذلك فإن جدولاً لأوامر التوريد "orders" مثلاً لابد أن يحتوي على حقلي رقم العميل customer-ID, product-code. اللذين يقابلهما حقلا customer-ID في جدول العملاء، وحقل product-code في جدول السلع. ومن ثم تتم عملية الربط بالعلاقات بين كل حقلين متشابهين في الاسم والقيمة. وتساهم مثل هذه العلاقات في تحسين أداء مهام الاختزان والتكشيف، وخاصة البحث والاسترجاع. وتصنف قواعد البيانات العلائقية في فئة قواعد البيانات التفاعلية والديناميكية. وتأخذ قواعد البيانات العلائقية منطق العلاقات كما ورد بخصائصه في علم الرياضيات[1].

- إيجابيات قاعدة البيانات العلائقية :

وتعد قاعدة البيانات العلائقية أفضل الأشكال التي تستخدم على الخط المباشر، ذلك لما لها من ملامح إيجابية تساهم في تحسين أداء مهام البحث والاسترجاع عن بعد. وهذه الملامح هي[2] :

- ♦ **تنظيم مجموعات كبيرة من البيانات:** حيث تمتنع قواعد البيانات عن تكرار القيم المتشابهة للبيانات، سواء أكان ذلك في أسماء حقول التسجيلات أم كان في قيم محتوى كل حقل داخل تسجيلات البيانات.
- ♦ **تكوين العلاقات يمنع تكرار قيم البيانات :** إن احتواء الجداول المختلفة على خطوط التعريف المسماة بالمفاتيح يجعلها مترابطة فيما بينها، مما يؤدي بدوره إلى منع تكرار قيم البيانات حتى داخل الجداول المختلفة.

---

(١) Hillman, Dave. Relational databases , the author, 2005, cited at 25/6/2007, cited at_ http://dhillman.com/theplace/databoot/reldbs.htm

(٢) Scammell, Lorna. Design your own database, the author, 2003, cited at 25/6/2007, cited at http://www.campus.ncl.ac.uk/databases/design/design.htm

♦ **إمكانية تكوين الروابط بين قيم متعددة للبيانات :** حيث يمكن على سبيل المثال تكوين رابط بين عنوان كتاب ومؤلف واحد أو أكثر من مؤلف.

♦ **التحكم التام في تحديث البيانات :** عندما يتم تخزين البيانات في أماكن محددة فإن تحديث مجموعات البيانات لا يمكن أن يشوبه الخطأ في التعامل مع مجموعة دون الأخرى.

♦ **صحة عمليات الإدخال :** على غرار التحديث، فإن إدخال البيانات يتم في أماكن تسكين مجموعات البيانات ذات الصلة.

♦ **تأمين البيانات :** حيث لا يمكن أن تتأثر مجموعة من البيانات بمجموعة أخرى تم تعديلها أو حذفها.

- سلبيات قاعدة البيانات العلائقية :

غير أن الإيجابيات السابقة التي توفرها قواعد البيانات العلائقية، تقابلها بعض ملامح النقص في الأداء، ومنها ما يلي:

١- **تأثير بعض الحقول على أخرى:** إذا ما تمت عملية الربط بين الحقول فإنه يحذر تأثر القيم المترابطة فيما بينها. لذا فإن عمليات الحذف والإضافة تتم فقط من خلال مدير قاعدة البيانات.

٢- **تحتاج إلى مراجعة شاملة :** حيث لا بد أن يعقب عمليات الحذف عملية مراجعة على روابط وقيم الدوال المترابطة كاملة للتأكد من عدم تأثرها.

٣- **صعوبة التقارير والمراجعات :** ويأتي ذلك نتيجة لتسكين مجموعات البيانات في جداول متعددة مختلفة.

تختزن قواعد البيانات العلائقية مجموعاتها من البيانات في شكل وحدات كبيرة من البيانات هي الجداول، ثم تتدرج هذه الوحدات إلى وحدات أصغر فأصغر وصولاً إلى حقل البيانات وتمثيلات الحروف. ولعل أهم ما يميز النظم العلائقية في رأي الباحث شيئان هما:

أولاً : طبيعة بناء الجداول : يختص كل جدول داخل قاعدة البيانات بمجموعة محددة من البيانات التي تخص موضوع واحد ، ولا يمكن لقاعدة البيانات أن تحتوي على جدولين متشابهين في الموضوع. وتدور حقول الجدول بالتالي حول موضوع أو نشاط فرعي لمحتوى قاعدة البيانات. وعندما تصاغ حقول الجداول بشكل صحيح، فإن ذلك يمنع تكرار البيانات أو تداخلها بين الجداول. وتتميز جداول النظم العلائقية بخصائص مثل :

- تملك جداول النظم العلائقية قيمًا ذات دلالة ؛ لما لها من تفرد أو عدم تداخل مع أي قيم أخرى.
- لا تتكرر الصفوف أو التسجيلات فيما بينها ، أو مع صفوف وتسجيلات جداول أخرى.
- تتشابه قيم أعمدة الجدول لما لها من تكامل في التعبير عن موضوع واحد.
- التتابع أو الترتيب بين الأعمدة أو الحقول غير مهم أو مؤثر في تنظيم البيانات واسترجاعها.
- التتابع أو الترتيب بين الصفوف أو التسجيلات غير مهم أو مؤثر في تنظيم البيانات واسترجاعها.
- يحظى كل حقل أو عمود باسم متفرد غير متشابه مع الحقول الأخرى للجدول.

ثانيًا: العلاقات أو الروابط :

تعتبر العلاقات الصفة المميزة لقواعد البيانات العلائقية عن غيرها. وتكمن أهمية العلاقات في تحقيق الترابط بين جداول قاعدة البيانات ومكوناتها ، مما يؤثر بدوره على قدرات قاعدة البيانات في تنظيم مجموعات البيانات ومعالجتها واسترجاعها بكفاءة عالية. وتستخدم النظم العلائقية الحقول المتشابهة في جدولين أو أكثر للربط بينهما. وتتأثر طبيعة العلاقات بين الحقول بحسب طبيعة

الحقول ذاتها. فتختلف على سبيل المثال العلاقات بين الحقول ذات المفاتيح الأساسية عن تلك الحقول غير المعرفة. وتبعًا لذلك تخرج علاقات النظم العلائقية على ثلاثة أنواع أساسية هي :

١) علاقة رأس برأس.

٢) علاقة رأس بأطراف.

٣) علاقة أطراف بأطراف.

قواعد البيانات العلائقية متعددة الوسائط ORDBMS:

يعبر هذا المصطلح عن نظم البيانات العلائقية متعددة الوسائط Object/relational database management systems. ويعد هذا الشكل وليدًا للنظم العلائقية ؛ حيث يرث مختلف مهام وطبيعة الأداء للنظم العلائقية، غير أنه يختص بأشكال من البيانات أكثر تعقيدًا من البيانات النصية مثل ؛ مجموعات الوسائط المتعددة كالصوت والصورة وغيرها من مستحدثات الأشكال الرقمية مثل X-rays ومقاطع الصوت Audio and video segments . وتضطلع هذه النظم بمهام أكبر من ناحية تنظيم ومعالجة وتحليل الوسائط المتعددة، فضلاً عن تدابير البحث والاسترجاع والعرض على شاشات تلائم طبيعة الوحدات المستدعاة.

ولكي تؤدي هذه النظم مهامها في معالجة وبحث الوسائط المتعددة، فإنها تعتمد على استخدام تقنيات عمل مختلفة مثل : استخدام لغات التعريف بالبيانات data definition languages (DDLs) حتى يمكنها التعامل مع نماذج متنوعة من البيانات في وقت واحد، كما تستخدم تقنية SQL3 في أداء مهام البحث وإنتاج شاشات عرض النتائج بما يلائم قواعد البيانات ODBC, JDBC. وتنتج هذه النظم بواسطة شركات مثل IBM, Inform ix, and Oracle[1].

(١) Multimedia database management systems.ppt Cited 30/6/2007, cited at http://ieeexplore.ieee.org/xpl/RecentCon.jsp?punumber=3975

الشكل الثاني : قواعد البيانات الهرمية Hierarchical DBMS:

تختزن قواعد البيانات الهرمية مجموعات البيانات في شكل البنية الشجرية أو الهرمية. وتنطلق في ذلك معتمدة على التواصل بين الآباء والأبناء في مجموعات البيانات. وعلى عكس النظم العلائقية فإن قواعد البيانات الهرمية تحتوي على قيم بيانات متكررة أو متداخلة ؛ حيث تنطلق في تسجيل البيانات في شكل سلاسل من التسجيلات المكونة من حقول متكاملة متشابهة الخصائص، ويأخذ كل فرع من السلسلة خصائص الفروع الأعم وصولاً إلى التسجيلة الأعلى. وتشبه العلاقات بين التسجيلات في مستويات التقسيم والبناء المختلفة، طبيعة العلاقات نفسها بين الجداول في النظم العلائقية المعتمدة على علاقات علم الرياضيات. ولإنشاء علاقة في البنية الهرمية فإن قاعدة البيانات تستخدم تواصل الأب بالابن حيث يمكن إنشاء عدد لا نهائي من العلاقات بين التسجيلة الأم والتسجيلات المنبثقة عنها في النوع نفسه، وتمثل العلاقة بـ(**N : 1**). ومثال لذلك فإن مؤسسة ما قد تلجأ لتسجيل بيانات موظفيها داخل قاعدة البيانات، ومن ثم تحتاج إلى تسجيل بيانات أبناء هؤلاء الموظفين، وعلى ذلك فإن كل موظف تنتج عنه تسجيلات تتحدد بعدد أولاده. وعلى ذلك فإن النوع الأوحد للعلاقات داخل النظام الهرمي هو علاقة رأس بأطراف. لقد ظل استخدام هذا الشكل من نظم قواعد البيانات شائعًا في الفترة من بداية الستينيات حتى السبعينيات، وقد أنتج هذا النظام بواسطة IBM.

الشكل الثالث : قواعد البيانات الشبكية Network DBMS:

تزامن انتشار النظام الشبكي لقواعد البيانات مع شيوع استخدام النظام الهرمي. وقد جاء النظام الشبكي لاختزان ومعالجة مجموعات البيانات نتيجة لأن هناك أنواعًا من مجموعات البيانات لا تربطها علاقات الأب والابن، وإنما تقف جميعا على درجة واحدة من التساوي بجانب ارتباطها معًا بخصائص مشتركة.

ومن ثم فإن قواعد البيانات الشبكية تعتمد بالأساس على نوع واحد من العلاقات هو علاقة أطراف بأطراف، وذلك خلافًا لنوع رأس بأطراف الذي يعتمد عليه النظام الهرمي. وفي عام ١٩٧١ عرف مؤتمر لغات نظم البيانات Conference on Data Systems Languages (CODASYL) النظام الشبكي على أنه مجموعة مترابطة من مجموعات التسجيلات. وتحتوي كل مجموعة على خصائص مشتركة بين مختلف التسجيلات تجعل هناك تواصلاً بإنشاء العلاقات بين هذه التسجيلات. إلى جانب ذلك تملك واحدة أو أكثر من تسجيلات مجموعة ما بعض الخصائص المتشابهة مع تسجيلات مجموعة أخرى، ومن ثم يتم الربط بينهما. وقد تملك تسجيلة واحدة خصائص تسمح لها بالربط مع أكثر من مجموعة من التسجيلات. وبتصور هذا العدد غير المحدود من تسجيلات البيانات والعلاقات الرابطة بينها، يمكن تخيل شكل ذي طابع شبكي يجمع بين التسجيلات والعلاقات. ويطلق على التسجيلات في هذه الحالة نقاط الربط أو التشابك.

وتتشابه أشكال نظم إدارة البيانات العلائقية والهرمية والشبكية في خصائص مثل : إمكانية معالجة وتنظيم قدر كبير من مجموعات البيانات، وإمكانية التعامل مع خصائص بعض البيانات غير النصية. غير أن الباحث يرشح قواعد البيانات العلائقية للتعامل مع محركات البحث في بيئة الويب، وذلك لما لها من خصائص تنظيمية ومعالجة تلائم البحث والاسترجاع في بيئة الويب.

٤/٢ مكونات محتوى قواعد البيانات :

تهدف قاعدة البيانات الى إجراء مجموعة من العمليات على ما تختزنه من مجموعات البيانات حتى تغطي احتياجات مستخدميها من المعلومات. وعلى ذلك تعتبر قاعدة البيانات في الأساس مستودعًا لاختزان كمية من البيانات يتم معالجتها. وتقسم مجموعات البيانات على اختلاف حجمها إلى مجموعة من

وحدات تخزين البيانات، تكون في مجملها البنية الهرمية لتخزين البيانات. ولعل أبرز تشبيه لقاعدة البيانات هو المبنى السكني المكون من مجموعة من الطوابق التي تقابل الجداول والمكونة بطبيعتها من مجموعة الوحدات السكنية التي تقابل التسجيلات، ثم الغرف التي تقابل حقول التسجيلة. ويظهر التقسيم التالي تدرج وحدات التخزين من الأكبر الى الأصغر حجمًا(١):

١/٤/٢ الملف أو الوحدة الكاملة File :

يطلق اسم الملف في أحيان كثيرة على وحدات الجداول داخل قاعدة البيانات، وفي أحيان أخرى يستخدم هذا المصطلح للتعبير عن قاعدة البيانات كوحدة واحدة. كما أن مصطلح ملف البيانات شائع الاستخدام في البيئة الرقمية ويحدد مفهومه في إطار بيانات نصية من وثيقة word يطلق عليها اسم ملف نصي، أو مادة مصورة رقمية يطلق عليها ملف مرئي أو أي نوع آخر من بيانات pdf , ppt... وغيرها من الملفات الموجودة على الويب.

٢/٤/٢ الجداول Tables : يحتوي كل منها على مجموعة من تسجيلات البيانات حول موضوع أو نشاط محدد. ويتألف الجدول كوحدة تخزين للبيانات من شقين أساسيين هما الأعمدة والصفوف ؛ تحدد الأعمدة بحقول التسجيلة الرقمية ويوضع على رأس كل عمود اسم الحقل المكون للجدول. وتبدأ تسجيلات الجدول من الصف الثاني ممتدة إلى تقاطع الأعمدة حتى تستقبل البيانات المدخلة وفقًا لكل حقل أو عمود. ويمثل الجدول التالي أحد جداول تخزين البيانات داخل إحدى قواعد البيانات، ويتكون من أعمدة تحدد الحقول، وصفوف تكون التسجيلة وخلايا تسجل بها البيانات.

(١) Parts of a Database. Thinkquest, 2001, cited at 25/6/2007, cited at http://library.thinkquest.org/C0115420/Cyber-club%20800x600/DatabaseS/Parts%20of%20a%20Database.htm

نموذج لجدول داخل قاعدة البيانات

الرقم العام	العنوان	المؤلف	الناشر	الطبعة
١	علم المعلومات	أحمد بدر	دار غريب	الأولى
٢	تكنولوجيا المعلومات	هانئ محيي الدين	الدار المصرية	الأولى
٣	محركات الصور	سيد ربيع سيد	مكتبة الملك فهد الوطنية	الأولى

٣/٤/٢ التسجيلات Records : وهي صفوف الجداول المكونة من اتحاد أعمدة الجدول في صف واحد لكل تسجيلة. وتسمى نقاط تقابل الأعمدة مع الصفوف بالخلايا التي تصب بها البيانات المنتمية لكل حقل من الحقول. وتشكل التسجيلة وحدة قائمة بذاتها تتجمع بها كل الحقول المكونة للجدول ؛ حيث يوضع بها مختلف البيانات المتكاملة عن سلعة أو كتاب أو خدمة... وغيرها من الوحدات التي تقوم عليها قواعد البيانات. ويمكن القول إن كل صفوف الجداول داخل قاعدة البيانات هي تسجيلات ماعدا الصف الأول الذي يحمل عناوين الأعمدة أو مسميات الحقول.

٤/٤/٢ الحقول Fields : وهي أعمدة الجدول التي تسجل بها البيانات. وتحدد عناوين الحقول بعناصر البيانات الواجب توافرها داخل كل تسجيلة ؛ فتسجيلات الكتب – مثلاً – تعنون حقولها بعناصر المؤلف والعنوان والناشر والطبعة... إلخ. وتعمل الحقول على إدخال أنواع مختلفة من البيانات داخل كل تسجيلة على حدة؛ حيث توضع البيانات النصية في حقول مثل الاسم والعنوان، وتوضع البيانات الرقمية في حقول مثل رقم الهاتف والرقم المعرف ID، كما توضع بيانات رسومية في حقول مثل التخطيط البياني أو الصور الشخصية... حتى تستكمل إدخال مختلف بيانات التسجيلة الواحدة. وإلى جانب تحديد نوع البيان، تحدد الحقول أيضا حجم البيانات المتوقع إدخالها داخل كل خلية من خلايا التسجيلة. وتضع بعض نظم إدارة قواعد البيانات خصائص كثيرة للتعامل مع حقول التسجيلة، وهو ما يعكس أهمية بناء الحقول في التأثير على أداء قاعدة البيانات. ومن هذه الخصائص صفة مطلوب أو غير مطلوب عند إدخال البيانات، وهي الخاصية التي

تحتم إدخال بيانات الحقل للسماح بإدخال التسجيلة كاملة ؛ أيضًا خاصية مفهرس أو غير مفهرس التي تحدد خصائص التكشيف أو عدم التكشيف للبحث بهذا الحقل.

٥/٤/٢ التمثيلات Characters : وهي الحروف والرموز المسجلة وتقاس بوحدة التخزين البت bits. وهذه الوحدة هي أساس عمل قاعدة البيانات كاملة، وبدونها فلا مكان للحديث عن مهام أو نظم قواعد البيانات. وتعرف البيانات بأنها أصغر بت bit يمكن بمجموعة منه تمثيل نص أو صورة أو صوت أو أي وسيط رقمي يتم اختزانه في بيئة التخزين الرقمية. وتحدد أنواع البيانات التي توجد داخل قواعد البيانات سواء النصية أو غير النصية بأنواع معيارية يمثلها الجدول (٢-٣):

الجدول (٢- ٣) الأنواع المعيارية للبيانات داخل قواعد البيانات

Number – integer and whole numbers
Strings – fixed and variable length (memo)
Date and time
Currency
Objects
Images
Digital audio and video
BLOBs (Binary Large Objects) – various forms of binary data

تضم قواعد بيانات الويب مختلف الأنواع السابقة من البيانات، وتؤدي قواعد البيانات مهامها بمعيار واحد، وهذا ما يجعلنا نستشف حقيقة مؤداها أن قواعد البيانات النصية أو الببليوجرافية أو حتى الوسائط المتعددة تعمل جميعًا بآلية عمل واحدة لمعالجة وتنظيم وبحث مجموعاتها الرقمية ؛ حيث يتم تنظيم مجموعات من البيانات في شكل تسجيلات مكونة من مجموعة من الحقول، وترتبط هذه التسجيلات بأشكال ربط هرمية أو شبكية أو علائقية من خلال الجداول التي

تكونها معًا، ثم البحث والاستدعاء بالاعتماد على تنشيط كل أو بعض هذه الحقول أثناء التكشيف والمضاهاة. وتحدد لغة البحث باستخدام تيجان SQL أو معايير عالمية لبحث المعلومات مثل Z39.50، غير أن الاختلاف بين قواعد البيانات يكمن فقط في متطلبات معالجة بعض أشكال البيانات مثل تحليل محتوى المواد المسموعة والمرئية أو الأشكال المعقدة للبيانات. وسيتم بناء النظرية النهائية لنظم استرجاع قواعد البيانات على الويب في هذه الدراسة على آلية أو معيار العمل الموحد لقواعد بيانات الويب.

٥/٢ استراتيجيات بحث قواعد البيانات على الويب:

تختلف قواعد بيانات الويب عن غيرها من مصادر الويب؛ لأنها تعد مصدرًا لمجموعة من المصادر الرقمية. فتحوي قواعد بيانات الويب مجموعات من البيانات متمثلة في شكل ملفات نصية أو وسائط متعددة. وهناك فارق كبير بين البحث عن قواعد البيانات والبحث داخل قواعد البيانات. ومن ثم يركز هذا الجزء من الدراسة على سرد مختصر لجوانب بحث قواعد البيانات على الويب. وتبدأ هذه العناصر بالحديث عن الشق الأول : البحث عن قواعد البيانات، ثم الشق الثاني : بحث محتوى قواعد البيانات :

١/٥/٢ البحث عن قواعد بيانات الويب:

تعامل قاعدة البيانات في حالة البحث عنها كمصدر واحد للمعلومات الرقمية. والبحث عن قواعد البيانات يقف فقط عند مستوى قاعدة البيانات ولا يتعدى ذلك إلى محتوى قاعدة البيانات المكون من ملفات أو مجموعات البيانات. ويتم تكشيف قاعدة البيانات باستخدام مصطلح أو مصطلحين على الأكثر للدلالة على موضوع القاعدة كله. وتكون نتائج البحث هنا عبارة عن أسماء قواعد البيانات التي تخدم استفسار البحث. أي أن قاعدة البيانات ستكون نظام الاسترجاع في الشق التالي عند بحث محتوى قاعدة البيانات، أما في هذه الحالة

فإن أدوات الاسترجاع تعد كنظم لاسترجاع قواعد البيانات. وسوف يتطرق الباحث عند الحديث عن الصورة النهائية لنظم استرجاع قواعد البيانات إلى شكل ثالث هو استرجاع محتوى قواعد البيانات من خلال أدوات بحث الويب، وهو ما سيرد تفصيله ختام هذه الدراسة.

٢/٥/٢ بحث محتوى قواعد بيانات الويب:

تتكون قواعد البيانات من مجموعات البيانات التي يتم اختزانها بشكل محدد. وتعمل نظم إدارة قواعد البيانات على تنظيم البحث ومعالجته بتوفير آليات البحث والاستدعاء. وتسمى آليات البحث والاستدعاء باستراتيجيات البحث لملفات قواعد البيانات. وتعمل مختلف نظم قواعد البيانات على توفير مستويين من لغة الاسترجاع، هما :

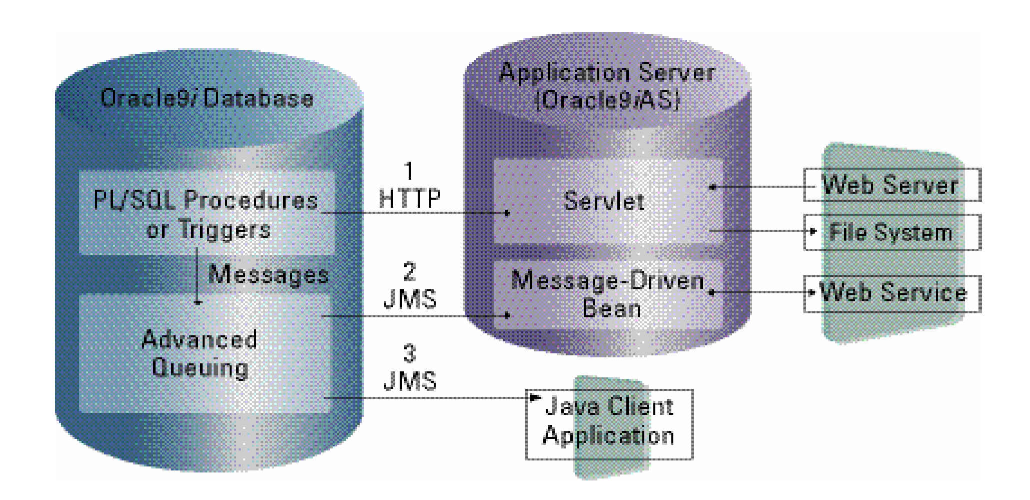
الأول : نظم الاختيار من متعدد ويسمى بأسلوب التصفح.

الثاني : هو الأمر باستدعاء وحدات محددة ويسمى بالبحث الحر أو لغة الأوامر. ولكن قبل التعرض لهذين المستويين يجب التعرف إلى آلية عمل نظم قواعد البيانات في تلقي الأسئلة وإرسال الإجابات في بيئة الويب.

١/٢/٥/٢ آلية تبادل الأسئلة والإجابات:

يمر سؤال البحث أو استفسار مستخدم قاعدة البيانات عبر ثلاث مراحل أساسية وصولاً إلى قاعدة البيانات. وبالمثل تتكرر هذه المراحل عند استدعاء النتائج. وتأتي هذه المراحل نتيجة لاختلاف بيئتي html الويب ونظام قاعدة البيانات، ومن ثم كان الاحتياج لبيئة ثالثة وسيطة بينهما متمثلة في لغة التحويل scripting language. ويعكس الشكل (٢-٧) آلية بحث قواعد البيانات في بيئة الويب[١].

(١) O'Rourke, Cameron. The active database. oracle , 2003, cited at 25/6/2007, cited at_ http://www.oracle.com/technology/index.html



الشكل (٢-٧) آلية بحث قواعد البيانات في بيئة الويب

يوضح الشكل (٢-٧) آلية بحث قواعد البيانات على الويب متمثلة في المراحل التالية:

١- يجري مستخدم الويب استعلامًا أو استفسارًا بلغة html ويرسله إلى خادم الويب.

٢- يرسل خادم الويب استفسار البحث الى وسيط العمل (لغات التحويل CGI) بين خادم الويب ونظام قاعدة البيانات.

٣- يستقبل نظام إدارة قاعدة البيانات استفسار البحث ثم يقوم بإجراء البحث واستدعاء النتائج لإرسالها الى وسيط التحويل، ومن ثم إرسالها الى خادم الويب أو متصفح الويب.

٢/٢/٥/٢ استراتيجيات بحث قواعد بيانات الويب:

قد تعتمد قواعد البيانات في بحث ملفاتها أحد أسلوبي التصفح والأوامر أو أن تستخدم الأسلوبين معًا.

◆ **أسلوب البحث بالتصفح:**

يلائـم هـذا الأسـلوب مـستخدمي قواعـد البيانـات ذات المـستوى المتواضـع في مهـارات البحـث. ويعتمـد التـصفح علـى مبـدأ تمثيـل مختلـف وحـدات البيانـات أو التسجيلات الداخلية برمز أو مثال يستطيع المستخدم من خلاله استدعاء وحدات موضوعية محددة. غير أن أسلوب التصفح يلائم فقط قواعد البيانات ذات الحجم الصغير، ذلك لكون تصفح حجم كبير من مجموعات البيانات قد يتسبب في تشتت المستخدم وضياع الاتجاه المخصص لبحث محتوى القاعدة. ويوجد شكلان أساسيان لتصفح قواعد البيانات هما :

أولاً: التـصفح بالمـصطلحات : وفي هـذا الـشكل تعتمـد قاعـدة البيانـات علـى مجموعة من المصطلحات المشتقة من التسجيلات أو المعينة للتعبير عن محتواها. وترتيب هذه المصطلحات فيما بينها من الأعم الى الأخص في فئات موضوعية مختلفة. وقد تعتمد قواعد البيانات على تقسيم موضوعي خاص بها لترتيب مصطلحات التصفح، أو أن تعتمد على قوائم رؤوس موضوعات مقننة معدة مسبقًا كقائمة عامة مثل قائمة مكتبة الكونجرس أو قائمة متخصصة في مجال قاعدة البيانات.

ثانيًا: التصفح بوحدات البيانات : يناسب هذا الشكل قواعد بيانات الوسائط المتعددة التي تحوي مجموعات الصور أو الأشكال الرسومية الأخرى. هنا تستخدم مجموعات اللقطات المصورة الممثلة لمجموعات البيانات الداخلية في عرض محتوى قاعدة البيانات، بدلاً من الاعتماد على المصطلحات. وبهذا الشكل يستطيع مستخدم قاعدة البيانات الانتقاء بين وحدات الصور قبل إجراء عملية البحث.

◆ **أسلوب البحث بالأوامر:**

أولاً: البحث بالحقول:

يقصد بالبحث بالحقول هنا الاعتماد على رموز حقول الوصف الببليوجرافي في استرجاع وحدات من المعلومات تشترك في مجموعة من الحقول مثل تاريخ نشر

محدد أو مؤلف واحد أو موضوع مخصص. وتأتي آلية البحث بحقول الوصف تبعًا لاعتماد قواعد البيانات على وضع تسجيلات ببليوجرافية لما تحتويه من مجموعات البيانات، سواء أكانت هذه المصادر صورًا أو نصوصًا في أشكال مختلفة. وقد تستخدم قاعدة البيانات القوائم المنسدلة للاختيار بين حقول تسجيلات البيانات، أو تمكين مستخدم الويب من إدخال حقول البحث بلغة الأوامر.

ثانيًا: البحث بالكلمات المفتاحية:

يعتمد البحث بالكلمات المفتاحية على طرح مجموعة الكلمات الدالة على موضوع مستخدم قاعدة البيانات. وتتبع قواعد البيانات مجموعة من آليات ومعاملات البحث للربط بين كلمات البحث بحيث يمكن لها نهاية تشكيل ما يسمى باستراتيجية البحث. ومن أهم آليات البحث استخدام المنطق البولياني bolean logic المعتمد على معاملات البحث and , or , not.

٦/٢ خاتمة:

تتسم قواعد بيانات الويب بخصائص تتفرد بها عن مصادر الويب الأخرى من حيث المهام وتخزين البيانات وبحثها. ولاختلاف موضوعاتها وأنواع البيانات التي تحملها الى جانب طرق تصميمها، فإن هناك تقسيمات عدة تتنوع على أثرها قواعد البيانات. وتؤدي قواعد البيانات مهامها من خلال نظم إدارة قواعد البيانات التي تتباين في قدراتها وأدائها فضلاً عن اختلافها من ناحية التصميم وبنية تخزين البيانات؛ حيث توجد النظم العلائقية والهرمية والشبكية. ومرحلة البحث فإن من المهم تحديد اتجاه بحث قاعدة البيانات هل هو بحث للمحتوى أو بحث عن قاعدة البيانات ذاتها.

الفصل الثالث

أدوات تنظيم وبحث الويب غير المرئية

٣ / ٠ تمهيد :

يتناول هذا الفصل جوانب تنظيم واسترجاع قواعد بيانات الويب كمصادر للمعلومات الرقمية ، من خلال أداء أدوات بحث الويب بنوعيها المحركات والأدلة. وبجانب ذلك سيتم التعرض لطبيعة عمل محركات بحث الويب المتعددة ، لما لها من خصائص بحث قواعد بيانات محركات البحث. ويحاول الباحث بناء فكرة تطوير محركات بحث الويب الى محركات بحث متعددة ، غير أن التعددية هنا لا تشمل البحث في محركات البحث ، وإنما البحث داخل محتوى قواعد بيانات الويب غير المرئية ، بجانب ذلك يتطرق الفصل الى دراسة معايير البحث عن بعد التي تستخدم لتحقيق التوافق بين نظم الاسترجاع بعضها بين بعض. كما يركز هذا الفصل بشكل كبير على خصائص عمل برنامج الزاحف والآليات التي يمكن التحكم بها لتهيئته للعمل مع قواعد البيانات. ومن خلال هذا التمهيد يتجلى هدف الفصل الثالث في التعرف إلى مدى إمكانية عمل محرك بحث متعدد لقواعد البيانات ذات برنامج زاحف متوافق مع بنية قواعد الويب غير المرئية.

١/٣ تكشيف قواعد بيانات الويب واسترجاعها :

تستخدم كلمة تكشيف مع قواعد البيانات بشكلين مختلفين ؛ الأول : ما تعنيه كلمة تكشيف من دلالة على معالجة وتنظيم تسجيلات قاعدة البيانات ، فتخصص قاعدة البيانات حقولاً نشطة يمكن من خلالها تكشيف وحدات البيانات المختزنة داخل مستودع قاعدة البيانات. ومن ثم يكون التكشيف أساس عملية التنظيم والاسترجاع لمحتوى قواعد البيانات. أما الاستخدام الثاني الذي يعنينا هنا فهو تكشيف قاعدة بيانات الويب غير المرئية ، والتعرف إلى محتواها ورصد طبيعة موضوعاتها ، وهو ما يساهم في إخراجها من دائرة الظلام إلى

برامج الزاحف داخل محركات البحث. وتحدد مشكلة قواعد البيانات بالمسارات التالية:

١/١/٣ تكشيف قواعد البيانات ووصفها :

تشترك قواعد بيانات الويب مع مواقع الويب غير المرئية في مشكلة تخفي المحتوى. غير أن قواعد البيانات تمثل مشكلة ذات عمق أكبر في كونها تعتمد على نظم لإدارة مهامها، وقد صممت تلك النظم بشكل مغاير تمامًا لما توجد عليه الويب. وتتصف الويب غير المرئية بشقيها المواقع وقواعد البيانات بصعوبات تجعل من عملية الوصول إليها وتكشيفها مسألة صعبة أمام برامج زاحف محركات البحث، ومن هذه الصعوبات:

١- تصمم قواعد البيانات ببرامج متخصصة لإدارة قواعد البيانات، مما يجعل من صفحاتها بنية غنية بمعاملات ترتبط بعرض وحدات البيانات المستدعاة دون الاهتمام بفكرة تنظيم صفحات الويب وعلاقتها بمحركات البحث.

٢- تختلف بنية مسارات صفحات النتائج داخل قواعد البيانات عنها في صفحات html، لاحتوائها على تيجان تمثل معاملات الاسترجاع والترتيب لنتائج قواعد البيانات.

٣- تختلف قواعد البيانات في مجموعاتها كمًا وكيفًا عن الملفات والبيانات التي توجد داخل صفحات الويب المرئية، مما يعكس بدوره الاحتياج لأداء مختلف في التكشيف والاسترجاع لهذه القواعد.

وقد طرحت بعض الدراسات طرائق تكشيف محتوى مواقع الويب. من بين هذه الطرائق استخدام برامج الزاحف أسلوبها المعتاد حتى تصل إلى بعض صفحات الويب التي تحمل بنية خاصة. هذه البنية تختلف عن صفحة الويب في أنها تحمل فراغات إدخال للبيانات أو فراغات للبحث والارتباط بالموقع. ومن ثم تصنف

برامج الزاحف هذه الصفحات في فئة الويب غير المرئية. وتبدأ إجراءات تكشيف هذه المواقع باتباع ما يلي[1]:

- إرسال مجموعة من كلمات البحث المعدة مسبقاً إلى موقع الويب.
- استقبال مجموعة نتائج الاستفسار ثم تحليلها للتعرف إلى طبيعة الموضوع العام.
- يمكن التعرف إلى محتوى مواقع الويب غير المرئية من خلال النتائج الصفرية التي تعني أن كلمات البحث ذات دلالة سلبية.
- تستدعي محركات البحث هذه الصفحات أمام مستخدمي الويب في حالة البحث عن مواقع ترتبط بالكلمات نفسها الدالة المحددة لطبيعة وموضوع موقع الويب.
- تعمل النتائج المستدعاة أمام مستخدمي الويب كرابط بين محرك البحث وبين موقع الويب ليبدأ المستخدم بالتعامل المباشر مع الموقع.

أما الطريقة الأخرى لتكشيف مواقع الويب غير المرئية فتتعدى ذلك إلى مستوى قيام محركات البحث بالبحث اللحظي المباشر في مواقع الويب غير المرئية real time search، وذلك من خلال نقل استراتيجية بحث مستخدم الويب إلى الموقع ومن ثم استدعاء النتائج المطابقة لكلمات البحث.

يمكن لأدوات بحث قواعد البيانات استخدام الطريقتين السابقتين لتسكين قواعد بيانات الويب في تقسيم موضوعي يتدرج من الأعم الأعلى إلى المتخصص الأدنى في التقسيم. ويمكن أن تختزن تسجيلات ببليوجرافية تحمل كل تسجيلة منها حقول وصف تصف عنوان ومحتوى وموضوع وشكل المعلومات التي تحملها

---

(١) Ru, Yanbo and Horowitz, Ellis. Indexing the invisible web: a survey, University of Southern California, 2005, cited at 20/7/20007, citead at http://libraries.ou.edu/eresources/reserves/readings/7_31_2006_6_40_27_PM.pdf

قاعدة البيانات؛ لأن الغرض من تكشيف قواعد بيانات الويب هنا ليس تحديد موضوع كل تسجيلة داخل قاعدة البيانات، وإنما تحديد الكلمات الدالة على محتوى قاعدة بيانات الويب كله، ومن ثم تعد كل قاعدة وحدة تختزنها أداة بحث الويب غير المرئية لأجل استخدامها مرة أخرى عند تطابقها مع احتياجات المستفيد الموضوعية. وتسلك أدوات بحث الويب غير المرئية في تكشيف قواعد البيانات أساليب مختلفة، مثل اعتمادها على برامج تحليل محتوى قواعد البيانات من خلال طرح استفسارات البحث وتحليل النتائج المستدعاة، أو الاعتماد على العنصر البشري في دراسة محتوى واتجاهات قاعدة البيانات الموضوعية كما يحدث في أدلة البحث. والجدير بالذكر هنا أن ثمة مستويين لتكشيف قواعد البيانات خاصة ومواقع الويب غير المرئية عامة، ويتمثلان في : المستوى الأول : اعتماد أدوات البحث على تكشيف المحتوى غير المرئي من خلال واجهات البحث باستخدام الكلمات المفتاحية meta key word أو نماذج البحث داخلها page forms أو البيانات الخلفية metadata. أما المستوى الثاني فيتعدى ذلك إلى تحليل مجموعات البيانات الداخلية والتعرف إلى الكلمات المفتاحية الأكثر تواترًا الدالة على المحتوى العام لموقع الويب. وتستعين أدوات بحث الويب غير المرئية بمجموعة من العوامل التي تزيد من كفاءة التكشيف كمًا وكيفًا للويب غير المرئية. وهي[1]:

- تحليل شامل لصفحات وواجهات البحث للخروج بخريطة أو رسم sitemap لموقع الويب.
- التعمق في المحتويات المغايرة لبنية الويب مثل ملفات pdf, word, excel, videos, music.

(١) Search engine indexing for deep web pages. accuraCast limited, 2007, cited at 1/8/2007, cited at http://www.accuracast.com/seo-weekly/invisible-web.php

- تتبـع مـسارات اختـزان البيانـات داخـل قواعـد البيانـات الموزعـة أو المواقـع المترابطة.
- اختـزان المعلومـات الـواردة مـن تحليـل الويـب غـير المرئيـة في أدلـة التعريـف بقواعد البيانات داخل أداة البحث Bookmarks.

إن آخر ما يذكر عن تكشيف الويب غير المرئية هو فوائد هذه العملية التي تتمثل في :

١- دعم العمليات التعليمية في الوصول الى ملفات العروض التقديمية والملفات ذات الأشكال الأخرى التي لم تكن تقدمها محركات الويب قبل ذلك.

٢- ضـم نمـاذج مـن ملفـات الويـب ومـصادر معلومـات غابـت كـثيرًا عـن أدوات البحث، مثل المواد الرسومية والرسوم البيانية والصور والخرائط ... إلخ.

٣- الوصول الى المواد والمصادر البحثية ذات القيمة العالية.

٤- الوصول الى المواد النصية الشارحة للموضوعات مع نماذجها غير النصية الموضحة لموضوع البحث.

٥- تسمح بطرح المعلومـات العلميـة والبحثيـة في أشـكال متعـددة دون الاعتمـاد على صفحات الويب فقط.

٦- ارتفاع درجة إتاحة مصادر المعلومات أمام مستخدم الويب، خاصة تلك التي لم يكـن باسـتطاعته الوصـول إليهـا، الى جانـب بحـث صـفحات الويـب المتاحة.

وتواجـه أدوات بحـث الويـب غـير المرئيـة صـعوبات في التكـشيف باسـتخدام الترتيب الهرمي لموضوعات قواعد البيانات؛ حيث يفتقد الترتيب الهرمـي الدلالـة أو المعنـى لـبعض تجمعـات قواعـد البيانـات، ممـا يـضع حاجـة ملحـة للاسـتعانة بمـصطلحات ذات دلالـة ثابتـة واضـحة. وعملـت أدوات بحـث مثـل Vivisimo and Quigo على استخدام آليات تكشيف أخرى مثل اقتباس المعلومات Information Extraction (IE) مـن واجهـات البحـث أو صـفحات مواقـع الويـب، ويفيـد ذلـك في

إعادة هيكلة معلومات موقع الويب دون المساس بخريطة الموقع أو تفاصيله site map. ويتم ذلك على مرحلتين في التكشيف هما[1]:

١- النتائج المصنفة موضوعيًا Categorized result : حيث تتبع أدوات البحث نفس أسلوب مستخدمي الويب في التعامل مع النتائج المسترجعة الناتجة عن استفساره البحثي، وذلك باختيار أدوات البحث لنتائج محددة، ثم تبدأ في تصنيف هذه النتائج بحسب درجة الصلة بموضوعات البحث. ومثال لذلك إذا ما كانت كلمة البحث المستخدمة هي "الرياضة" فإن اختيار المستخدم لوحدات النتائج يعتمد على مدى صلتها بالموضوع. ومن ثم تكون الوحدات المنتقاة هي الوحدات الدالة على موضوع الرياضة وتصنف تحت هذا المصطلح.

البحث المترابط Associative search : تستطيع أدوات بحث الويب غير المرئية ابتداع روابط بين مواقع الويب المتشابهة في المصطلحات الدالة على الموضوع أو المشتركة لفظيًا، وذلك بتكوين روابط بين الكلمات المشتركة لفظاً أو معنىً. ومن ثم يستطيع الباحث التحرك داخل قاعدة بيانات أدوات البحث معتمدًا على شبكة الروابط بين المواقع المترابطة.

تكشيف قواعد بيانات الوسائط المتعددة:

تحتاج قواعد بيانات الويب غير المرئية الى أنواع مختلفة من التكشيف، لاحتوائها على محتوى متنوع بين الوسيط النصي والوسائط المتعددة. ما سبق ذكره عن التكشيف يوافق بالدرجة الأولى التكشيف النصي TBI. أما المحتوى المصور متحركًا كان أو ثابتًا فإن ما يلائمه هو استخدام التكشيف المبني على المحتوى Content based indexing CBI. الذي يستخدم تقنيات تحليل الصور الرقمية وتحليل بنية ولون وأبعاد الصورة، وهذه التقنيات تستخدم دائمًا مع نظم استرجاع الوسائط المتعددة. وبالطبع هنا فإن أدوات بحث الويب التي ستعمل

(١) Web Ignite corporation indexing deep web content, Visibility Factor, 2002, cited 1/8/2007, cited at http://www.visibilityfactor.com/articles/indexing_the_deep_web.html

كوسيط بحث بين قواعد البيانات وبين مستخدمي الويب لابد أن تحتوي على آليات طرح استفسارات البحث على شكل صور أو أجزاء مصورة وليس في شكل كلمات أو عبارات بحث.

ويمكن القول إن واقع استرجاع الوسائط المتعددة داخل قواعد البيانات يختلف عنه في صفحات الويب. وذلك يرجع الى أن ملفات الوسائط المتعددة محاطة بعناصر أو حقول نصية مكونة التسجيلات داخل قواعد البيانات، مما يعني حتمية استخدام النص في البحث، وهذا يقلل بدوره من أهمية الاعتماد على البحث بالمحتوى أو النماذج المصورة. وهو خلاف ما تكون عليه ملفات الوسائط داخل صفحات الويب من فقدان النصوص أو انتفائها وحقول وصف لملف الصورة، مما يزيد بدوره من الحاجة للاعتماد على تقنيات البحث بالنماذج المصورة.

٢/١/٣ استرجاع قواعد البيانات كمصدر للمعلومات:

يتفق استرجاع قواعد البيانات كمصدر للمعلومات مع التناول السابق لعملية تكشيف قواعد البيانات، حيث يعتبر كل منهما قاعدة البيانات وحدة مفردة داخل مجموعة قواعد بيانات الويب غير المرئية. ودائمًا ما يكون أسلوب الاسترجاع انعكاسًا لأسلوب التنظيم. ويرى الباحث أنه يوجد اختياران فقط أمام محركات بحث الويب غير المرئية لاسترجاع محتوى قواعد البيانات، يوضحهما الباحث فيما يلي:

أولاً: محركات البحث كوسيط بحث لقواعد البيانات:

تقوم محركات بحث الويب غير المرئية هنا بالدور الذي تقوم به محركات البحث المتعددة ؛ من خلال استقبال استفسارات مستخدم الويب في شكل كلمات مفتاحية. ومن ثم تحدد محركات بحث الويب غير المرئية الاتجاه الموضوعي لكلمات البحث ومطابقتها مع قواعد البيانات لتحديد الوحدات ذات الصلة بالموضوع. ويتم ذلك اعتمادًا على عملية التكشيف السابق الحديث عنها

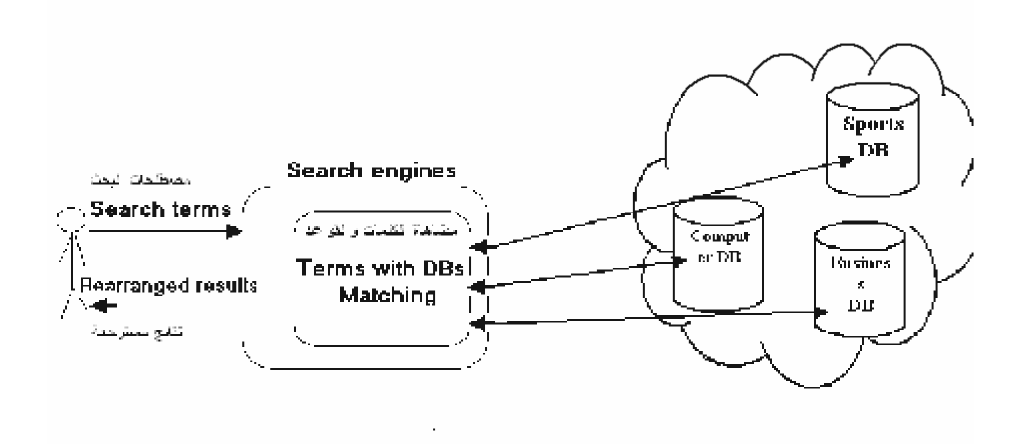
لقواعد بيانات الويب والتعرف إلى الاتجاهات الموضوعية لمجموعات البيانات بها. ثم تبدأ محركات البحث بإرسال استفسار البحث إلى قاعدة البيانات ذات العلاقة الموضوعية على أن تنتظر نتائج الاستفسار مرة أخرى، ثم إعادة هيكلتها وضبطها وفقاً لخصائص متصفح الويب وطبيعة وحدات البيانات المعروضة أمام مستخدم الويب، كما في الشكل (٣-١). وتقوم محركات البحث في هذه العملية بالإجراءات التالية:

١- تلقي استفسار البحث في شكل كلمات البحث واليات البحث المرتبطة بصياغتها.

٢- تحديد الوجهة الموضوعية لكلمات البحث ومطابقتها مع موضوعات قواعد البيانات المعرفة لدى محرك البحث.

٣- إرسال استراتيجية البحث بحسب لغة أو معيار البحث الملائم لقاعدة البيانات موضع البحث.

٤- تلقي نتائج البحث وضبطها بما يتوافق والعرض أمام المستخدم في بيئة html.



الشكل (٣-١) محركات البحث كوسيط لبحث قواعد البيانات

ثانيًا: محركات البحث كرابط لقواعد البيانات:

تقوم محركات البحث هنا بدور مختلف عن الدور السابق ؛ حيث تكتفي محركات بحث الويب غير المرئية بإطلاق برامج الزاحف لجمع مختلف قواعد البيانات على الويب. وتكشيف هذه القواعد بما يجعل لكل قاعدة بيانات تسجيلة خاصة بها داخل مستودع معلومات محرك البحث. ومن ثم يقوم محرك البحث بمضاهاة الكلمات البحثية الواردة في استراتيجية بحث مستخدم الويب وعرض تسجيلات قواعد البيانات كوحدات بين النتائج في شاشات عرض النتائج. ويمكن للمستخدم هنا الانطلاق من نتيجة عن قاعدة بيانات الى موقع هذه القاعدة لعمل بحث مباشر لمحتوى قاعدة البيانات وفق ما صُممت عليه، دون الوساطة التي كان يقوم بها محرك البحث. غير أنه يمكن القول هنا إن هذه العملية تعتمد بشكل أساسي على نشاط برنامج الزاحف وكفاءة عملية التكشيف لقواعد بيانات الويب غير المرئية.

ويمكن أن تتكون تسجيلات قواعد البيانات وعناصر الوصف بها كما يلي(١):

الجدول (٣-١) عناصر الوصف الفني لقواعد البيانات داخل محركات البحث

Database record fields	حقول تسجيلة قاعدة البيانات
Title	عنوان قاعدة البيانات.
Agency/Owner	المؤسسة أو المالك.
Database Abstract	مستخلص عن قاعدة البيانات.
Database Update Frequency	دورية وفترات التحديث لقاعدة البيانات.
Legal Constraints on Database Access	قيود الإتاحة الشرعية للوصول لقاعدة البيانات.
Form(s) of Digital Copies of Database	الأشكال الرقمية لقاعدة البيانات.
Database Digital Copy Media and Costs List	تكاليف النسخ أو مقابل التحميل.

(١) Google and the deep web. Blogger, 2007, cited 20/7/2007, cited at http//glinden.blogspot.com/2007/03/google-and-deep-web.html

تكاليف إعادة الإنتاج لقاعدة البيانات.	Report Reproduction Cost List
تقرير مقابل الحصول على خدمات قاعدة البيانات.	Custom Services Costs
قائمة نماذج البيانات داخل قاعدة البيانات.	List of Database Data Modules
قائمة حقول التسجيلات داخل قاعدة البيانات	Annotated List of Fields (one list per data module)

يعكس الجدول (٣-١) عناصر أو حقول الوصف الببليوجرافي لقواعد البيانات في شكل تسجيلات داخل مستودعات قواعد البيانات. وما على محرك البحث سوى إحلال نوع من التوافق بين موضوع كلمات بحث مستخدم الويب وبين التخصص الموضوعي لقواعد البيانات في اختيار الأنسب للبحث.

٣/١/٣ استرجاع قواعد البيانات كنظام استرجاع:

تختلف دلالة مصطلح الاسترجاع مع قواعد البيانات بحسب مستوى الاسترجاع ؛ فإذا كانت قاعدة البيانات تعامل كمصدر أو وحدة واحدة، عندئذ ستصبح نتيجة ضمن وحدات الاسترجاع في شاشة عرض النتائج لنظام الاسترجاع. أما إذا كانت قاعدة البيانات تعامل كنظام للاسترجاع وليس كمصدر، كان الاسترجاع لتسجيلات البيانات المختزنة بها، كما هو الحال في السرد التالي. ويعد البحث والاسترجاع أحد مهام نظام إدارة قاعدة البيانات، حيث يوفر النظام آليات للبحث تتكون من معاملات البحث التي تتوافق وطبيعة التسجيلات الداخلية. هذا فضلاً عن تحديد الأنماط العامة المتبعة في بحث التسجيلات مثل نمط التصفح أو البحث الحر. وعند هذا المستوى من الاسترجاع، تقوم نظم إدارة قواعد البيانات بتكشيف وتنظيم مجموعات التسجيلات الداخلية وتوفير الحقول النشطة للبحث.

وتعتمد مختلف نظم استرجاع محتوى قواعد البيانات على آليات محددة لبحث مجموعات البيانات، وإن اختلفت في درجات استخدامها. ويعرض الجدول (٣- ٢) أهم آليات بحث محتوى قواعد البيانات المستخدمة مع الكلمات المفتاحية[١]:

(١) سيد ربيع سيد. محركات بحث الصور الثابتة.- ط١.- الرياض: مكتبة الملك فهد الوطنية، ٢٠٠٧م.

الجدول (٣- ٢) أهم آليات بحث محتوى قواعد البيانات

آليات البحث بالكلمة الواحدة	آليات البحث بكلمات متعددة
البحث بالتطابق	البحث بالتقارب
البحث غير المطابق (البحث باللواصق)	استخدام المنطق البوليني
البحث الحر	طرح سؤال للبحث: قائمة توقف عن الأحرف والأدوات
البحث بحساسية الحالة	استخدام علامات التطابق " "
	البحث المعقد أو المركب
	البحث بمثال (- /+ " " تقدير الأهمية)

تؤدي قواعد البيانات مهامها كنظم للاسترجاع اعتمادًا على عاملين أساسيين؛ العامل الأول : عامل الدقة والتحقيق في مجموعة وحدات البيانات المسترجعة ويقابل ذلك عامل الاضطراب الدال على الوحدات المسترجعة التي لا تتصل بالموضوع. أما العامل الثاني : فهو الاستدعاء لكل الوحدات التي تتصل بالموضوع داخل نظام الاسترجاع، ويقابل ذلك فقد بعض وحدات البيانات ذات الصلة بالموضوع. ولكي تنجح قواعد البيانات كنظم استرجاع في رفع درجة التحقيق وزيادة نسبة الاستدعاء يجب أن تعتمد على مجموعة آليات البحث التي تهذب عملية استرجاع البيانات وتصحح صياغة المستخدم لمجموعة كلمات البحث.

وقد تناولت دراسة علمية بمدرسة الاتصالات والمعلومات ودراسات المكتبات بجامعة روتجرز بالولايات المتحدة School of Communication, Information and Library Studies (SCILS) at Rutgers University في فصل من فصولها مختلف نماذج آليات استرجاع المعلومات الرقمية مقسمة على ثلاث فئات، وتنبثق منها آليات البحث التابعة لها، وهي كما يلي[1]:

(١) Information Retrieval Models. (SCILS) at Rutgers University, 2002, cited at 1/8/2007, cited at Cited at http://www.scils.rutgers.edu/~aspoerri/InfoCrystal/Ch_2.html

☒ آليات المنطق البولياني: **Boolean Retrieval**

- آلية المنطق البولياني المعياري **Standard Boolean** (معاملات + ، –).
- آلية تضييق وتوسيع نطاق البحث Narrowing and Broadening Techniques.
- آلية البحث البولياني الممتدة Extended Boolean (and, or, not).

☒ آليات البحث الإحصائية : **Statistical Model**

- آلية البحث بالمعاملات الرياضية Vector Space Model : وتعني وزن الكلمات المهمة في البحث والوثائق ثم تحديد الوحدات المسترجعة.
- آلية البحث الاحتمالية Probabilistic Model : تعني إعادة البحث بحسب الكلمات الواردة في اختيارات المستخدم من نتائج البحث المسترجعة.
- آلية التكشيف الدلالي Latent Semantic Indexing.
- آلية التقسيم الموضوعي للوثائق Document Clustering.

☒ آليات البحث المبنية على الدلالة والمعرفة: **Linguistic and Knowledge-based**

- آليات البحث بالروابط الدلالية DR-LINK Retrieval : وهي تهتم بالاسترجاع المبني على المفهوم ؛ حيث استخدام التحليل اللغوي لكلمات البحث ووحدات البيانات.

٢/٣ أدوات بحث قواعد البيانات على الويب:

يرى مستخدمو الويب صفحات ومواقع الويب بعيون أدوات بحث الويب (المحركات والأدلة). ومن ثم تتطلب أدوات بحث الويب التحلي ببعض الخصائص في الأداء حتى نحول الويب غير المرئية الى ويب مرئية. ويهتم هذا الجزء بدراسة مختلف خصائص عمل محركات وأدلة البحث التي تعنى بالتعامل مع قواعد البيانات إضافة ومعالجة وبحثًا. ويعمد الباحث هنا إلى إصباغ محركات البحث

بصبغة التعامل مع قواعد البيانات حتى يكون ذلك تمهيدًا لتحديد الملامح النهائية لمحركات بحث قواعد البيانات في نهاية هذه الدراسة.

١/٢/٣ محركات بحث قواعد البيانات:

١/١/٢/٣ مفهوم محرك البحث:

لقد ذكرت بعض مواقع الويب تعريفات عن محركات البحث منها ؛ " أن محرك البحث هو بصفة عامة أداة تمكن مستخدمي الويب من البحث بمصادرها ، وأنه قد يعتمد على البرامج أو الإنسان في أداء عمله"[١]، وهذا التعريف قد عمل على تلخيص مهمة محرك البحث وإبرازها من ناحية مستخدمي الويب دون التطرق إلى الإجراءات التي يقوم بها في تنفيذ هذه المهمة. وثمة تعريف آخر لمحرك البحث هو "أنه عبارة عن أداة لإيجاد المعلومات على الإنترنت، ويتكون من مكونات خمس هي: العنكبوت أو الروبوت، والمكشف، وقاعدة البيانات، وآليات البحث، وواجهة الويب[٢]". وقد اختلف هذا التعريف عن سابقه من ناحية التركيز على مكونات محرك البحث، وإن كان الباحث يرى أنها قد تقتصر فقط على ثلاثة مكونات هي الزاحف والمكشف والبحث. وعلى ما سبق من تعريفات يرى الباحث أن أفضل تعريف لمحركات البحث هو " أنها إحدى أدوات بحث استرجاع المعلومات على الويب، وتعتمد على ثلاثة برامج أساسية في ذلك هي الزاحف والمكشف وآليات البحث، بما يوفر القدرة على الإضافة والمعالجة والبحث"[٣].

وتختلف محركات بحث الويب غير المرئية عن محركات الويب المعروفة في أنها تتعامل مع مصادر معلومات رقمية ذات بنية تختلف وبنية html. وأهم هذه

(١) Sullivan , Danny . Search engine optimization & marketing glossary, Submita website, Inc , 2004 , cited at 12/6/2004 , cited at http://www.submitawebsite.com/seo_glossary.html

(٢) What is a Search engine? metamend.inc , 2004 , cited at 12/6/2004 , cited at http://www.metamend.com/search-engine-map.html

(٣) سيد ربيع سيد. مصدر سابق.

المصادر اختلافاً في البنية والتنظيم هي قواعد البيانات موضوع الحديث. ويمكن اقتراح تعريف إجرائي لمحركات بحث الويب غير المرئية على أنها "إحدى أدوات بحث واسترجاع الويب، التي تختص بتنظيم نوع من مصادر الويب ذات بنية مختلفة عن بنية النصوص المعيارية الفائقة HTML ، ويمثل هذا النوع من مصادر الويب في أغلبه قواعد البيانات، التي لا تستطيع محركات بحث الويب العادية الوصول إليها وتنظيمها أو استرجاعها".

٢/١/٢/٣ مهام وأداء محركات بحث الويب:

تعمل محركات بحث قواعد البيانات في الويب كنظم استرجاع ؛ من حيث مهام الإضافة والتزويد، والإعداد الفني والتنظيم، والبحث والاسترجاع، ويضطلع محرك بحث الويب كنظام استرجاع للويب بالمهام الآتية :

أولاً: الإضافة والاقتناء:

تعتمد محركات البحث في عملية الإضافة والتزويد على البرامج الآلية (الزاحف أو العنكبوت). وتختلف محركات البحث عن أدلة البحث في أن الأدلة البحثية تعتمد على الأفراد في مرحلة الإضافة والتزويد، وذلك عن طريق تصفح الإنترنت ثم تجميع المواقع التي وقع عليها الاختيار، وفهرستها وتكشيفها لتوضع ضمن قاعدة بيانات الدليل. وهذا يعني أن حجم التغطية في الأدلة البحثية ليست بالقدر الكافي لمواقع الويب المتنامية العدد. فالأدلة تعمل من منظور الأعداد القليلة التي يمكن التحكم فيها وإعدادها بما يحقق للباحث أكبر قدر من الدقة والتحقيق، بينما تعمل محركات البحث على تغطية واسعة للمواقع على الويب.

ومن ناحية أخرى فإن التزويد والإضافة للأدلة البحثية يكون عن طريق الطلبات التي تقدم من أصحاب المواقع ذاتهم، سواء كانت تجارية أو حكومية أو علمية أو شخصية أو غير ذلك. أما الإضافة والتزويد في محركات البحث يكون عن طريق الزاحف الذي يتحرك بين روابط الصفحات المختلفة ومن ثم تحليل كل

الروابط داخل صفحات الويب المرئية ثم زيارة هذه الروابط، إلا أن برامج الزاحف تختلف فيما بينها من حيث مستويات تحليل صفحات المواقع؛ حيث تقف بعض البرامج عند المستويين الأول والثاني لصفحات موقع الويب، وتتعدى بعض برامج الزاحف ذلك إلى المستويين الثالث والرابع. وهو ما يؤثر بشكل أساسي على أعداد صفحات الإضافة اليومية لبرنامج الزاحف. وهذا يترتب عليه تأخير في فترات التحديث. فالتزويد في محركات البحث يتم في فترات متقاربة كل ساعة أو كل يوم، وقد يكون في فترات متباعدة أسبوعيًا أو شهريًا. إلا أنه يمكن اعتبار هذه الفترات دورة زمنية يقوم فيها الزاحف بالعمل حتى العودة مرة أخرى لمثل هذه المواقع للتعرف إلى ما جد بها من تحديث، أو تغيرات في متن هذه الصفحات.

ثانيًا: المعالجة والتنظيم :

بعد أن تنتهي المرحلة الأولى بإضافة نسخة من موقع الويب إلى قاعدة بيانات محرك البحث، يبدأ عمل الجزء أو البرنامج الثاني في محركات البحث الذي يعرف بالمكشف أو المفهرس ، حيث يقوم بإعداد الكلمات الدالة أو عناصر الوصف لمواقع الويب حديثة الإضافة إلى قاعدة بيانات أداة البحث، وتعمل هذه المعالجة الفنية وفقًا لما ورد في المكتبات والمعلومات عن (التمثيل المكثف)، حيث لا يتم اختزان موقع الويب بأكمله في عدد من أدوات البحث، وإنما حقول وبيانات منها مسار هذا الموقع للربط به والوصول إليه، وكل ذلك في شكل تسجيلة كما هو الحال في الفهارس الآلية للمكتبات OPAC. وتتم عملية الفهرسة في حالة الأدلة البحثية بشكل يدوي، أما في حالة محركات البحث فإنها تتم آليًا، وفي حالة التكشيف والفهرسة الآلية فإن برامج التكشيف الآلي تستخدم بنية صفحة الويب المرئية في التعرف إليها ووضع كلمات أو مصطلحات معبرة عن الموضوع أو الموضوعات التي تضمها صفحة الويب؛ حيث يفحص البرنامج هنا كل أجزاء الصفحة التي تحتوي على كلمات تخص الموضوع من

كلمات العنوان، وكلمات النص، والكلمات المفتاحية التي وضعها مصمم صفحة الويب في رموز meta tags وتجمع هذه الرموز كل الكلمات التي يرى مصمم الصفحة أنها تعبر عن موضوع أو موضوعات صفحة الويب. وبعد تفحص كل الجوانب السابقة يتم وضع كلمات مفتاحيه داخل تسجيلة هذا الموقع أو وضع كل كلمات النص الطبيعية في وضع نشط ليتم مضاهاتها بعد ذلك مع كلمات البحث التي يضعها الباحث عند إجراء البحث. وتختلف مواضع التكشيف من برنامج إلى آخر؛ حيث تعمل بعض برامج التكشيف على الاكتفاء بالجزء الأول فقط من بنية اللغة المعيارية لصفحة الويب، كما تعمل بعض البرامج على تكشيف عناوين الصفحات بجانب مسار الصفحة، وتمتد قدرة برامج التكشيف في أحيان أخرى إلى تكشيف كامل متن صفحة الويب وهو ما يجعل محركات البحث في هذه الحالة تمتاز بوفرة في الاستدعاء مع انخفاض نسبة التحقيق.

ثالثًا: البحث والاسترجاع:

يعد الاستدعاء من أهم مقاييس كفاءة النظام. لكن الأمر لا يقتصر على النظر إلى النتائج بل يتعدى إلى ماهية التنظيم داخل نظام الاسترجاع؛ فمحركات البحث تعتمد على البحث الحر في كامل نص الموقع مع إمكانية البحث في العنوان والرموز المعيارية (metatags) وبذلك فإن الاستدعاء سوف يكون في أعلى درجاته، بينما يكون التحقيق في أقل حالاته؛ لأن آلية التكشيف التي تعتمد عليها محركات البحث هي استخدام اللغة الطبيعية، وهذا إنما يعكس بشكل جلي أن كل كلمات النص سوف تصبح كلمات مفتاحية يمكن من خلالها استدعاء الوثيقة (موقع الويب) اعتمادًا على مضاهاة الكلمات البحثية. أما الأدلة البحثية فإنها تعتمد في البحث على كلمات كشفية ومصطلحات يتم وضعها بالتدخل البشري تعبر بصدق عن الموضوع. وبالتالي فإنه يستبعد أية وثائق أو مواقع ليست ذات صلة، ويولد تحقيقًا عاليًا مع استدعاء منخفض.

ومن ناحية أخرى تكون النتائج في محركات البحث مرتبة حسب درجة الصلة بين الموقع والموضوع. ويختلف الوزن النسبي لموقع الويب وتقدير أهميته للموضوع إما اعتمادًا على عدد تكرار الكلمات المفتاحية التي ذكرت في السؤال، أو أماكن تواجد الكلمات المفتاحية والموقع بحيث إذا ما وردت في العنوان يكون للموقع أهمية أكثر من وجودها في النص نفسه. وبحسب الأهمية يأتي ترتيب النتائج بالموقع، فالأكثر أهمية للموضوع يأخذ الرقم (١) في الترتيب ثم ما يليه من المواقع على حسب هذا الوزن ... وهكذا. أما بالنسبة للأدلة البحثية فإنه عادة ما تحدد هذه الأهمية من قبل المفهرس نفسه الذي يحدد موضع الموقع في الترتيب.

مما لا شك فيه أن البحث في صفحات الويب غير المرئية يشكل صعوبة لمحركات البحث ؛ لأنها تتعامل مع شكل مختلف من معالجة المعلومات، إضافة إلى النماذج المتباينة لآليات البحث المستخدمة في قواعد بيانات الويب غير المرئية. وذلك بعكس ما هو عليه في الويب المرئية ؛ حيث تتم عملية البحث داخل مستودع المعلومات الخاص بمحرك البحث. وتستخدم محركات بحث الويب أسلوب التعامل مع الويب المرئية نفسه لتتعامل به مع الويب غير المرئية من إضافة ومعالجة وبحث، غير أن شكل كل مرحلة يختلف وفقًا للاختلاف بين صفحات الويب المرئية وقواعد بيانات الويب غير المرئية. وهو ما يتضح من محتوى العناصر التالية.

٣/١/٢/٣ محركات البحث وقواعد البيانات :

تختلف قواعد البيانات عن مواقع وصفحات الويب حجمًا وكيفًا في احتواء المعلومات الرقمية على الويب. ومن ثم لا تتعمد محركات البحث احتواء أو إضافة قاعدة البيانات ذاتها داخل مستودعاتها كما في حالة صفحات ومواقع الويب المرئية، وإنما تقيم محركات البحث رابطًا بينها وبين قاعدة البيانات، يتمثل في تسجيلة من حقول وصف محتوى وموضوع قاعدة البيانات؛ لاستحالة اختزان وتجميع قواعد بيانات الويب غير المرئية داخل مستودعات محركات البحث.

ويقوم محرك البحث بتحديد الموضوع العام لكلمات بحث المستفيد ، ثم تحديد قاعدة أو قواعد البيانات الملائمة لهذا الموضوع، ثم صياغة كلمات البحث بلغة البحث التي تعمل بها قاعدة البيانات لتوجيه استفسار البحث إليها وانتظار النتائج التي يتم تحويلها الى مستخدم الويب. ولكي تحقق محركات البحث نوعًا من التوافق بينها وبين قواعد البيانات فإنّ ثمة عددًا من التقنيات التي تحتاجها محركات البحث لذلك ؛ مثل قدرة برنامج الزاحف على تحديد واجهات البحث عند إضافة قواعد بيانات جديدة، أيضًا التوافق مع لغات البحث المستخدمة في قواعد البيانات[1].

١/٣/١/٢/٣ (الزاحف) الوصول والإضافة لقواعد البيانات:

يعرف برنامج الزاحف على أنه أحد البرامج العاملة في أداء مهام محركات بحث الويب، ويختص الزاحف بتزويد مستودع محرك البحث بمصادر الويب. ويؤدي الزاحف مهامه اعتمادًا على ملف داخله يسمى robot.txt ، يحمل مختلف التعليمات المحددة لعمل الزاحف وظيفةً وأداءً.

تكمن مشكلة استرجاع قواعد بيانات الويب غير المرئية في إمكانية الوصول إليها بواسطة برامج الزاحف أو العنكبوت. ويمكن القول إن (٩٠٪) من معاناة محركات بحث الويب في الوصول إلى قواعد البيانات تأتي من انتفاء وانعدام قدرة الزاحف على التوافق مع واجهات قواعد بيانات الويب. ومن ثم كان الجزء الأكبر من حل هذه المشكلة في تزويد برامج الزاحف بآليات العمل التي تمكنها من التلاقي وواجهات بحث قواعد البيانات، بل والقدرة على تحليلها ونسخها وإضافتها إلى مستودع المعلومات لمحركات البحث. وعندما كانت قواعد البيانات تعتمد في استرجاع محتوياتها على البحث المباشر من خلال واجهات

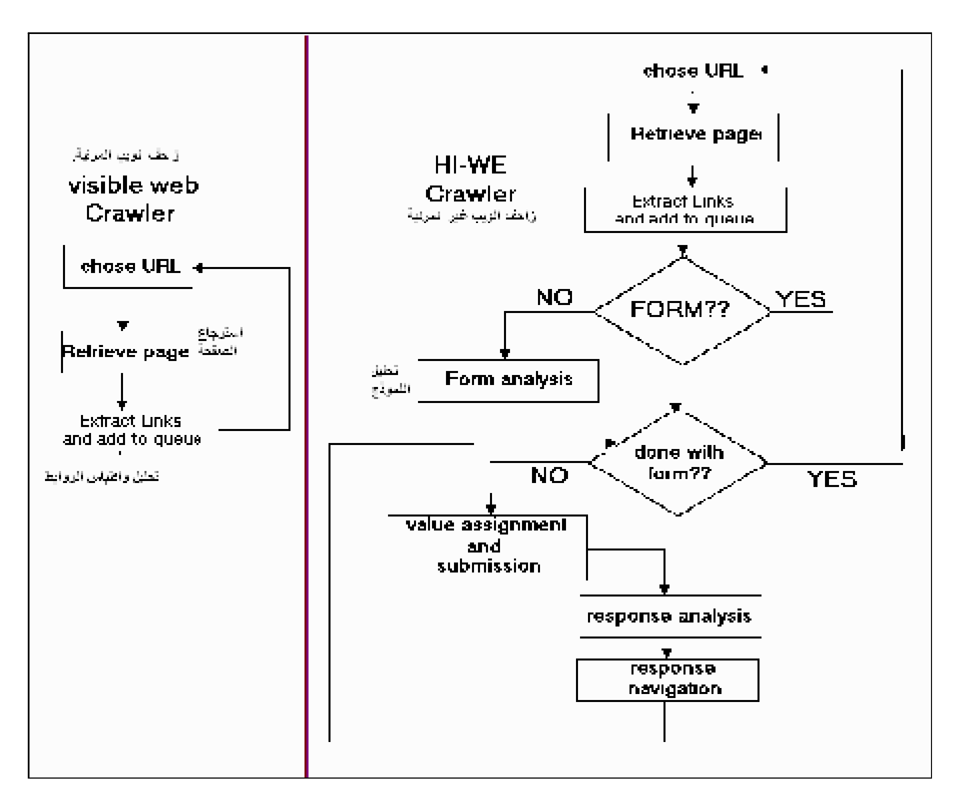
---

(١) Invisible web. CloserLook search services Inc. 2002, cited 5/8/2007, cited at http://www.searchengineguide.com/wi/2002/0327_wi2.html

البحث بها، لم يكن لبرامج الزاحف أو محرك البحث أي قدرة على التعامل مع هذه المصادر ذات المحتوى الضخم من ملفات المعلومات الرقمية. هذا، وقد ظلت برامج الزاحف تتبنى آلية واحدة للعمل منذ خروجه إلى فضاء الويب، ترتكز هذه الآلية على استخدام مسارات أو روابط صفحات الويب للوصول إليها، ومن ثم تعتمد في تحليل هذه الصفحات على تيجان صفحات الويب HTML metatags. وهذه الآلية بشكلها الحالي لا تتوافق مع واجهات بحث قواعد بيانات الويب التي تتكون في مجملها من نموذج لبحث قاعدة البيانات معتمدًا على حقول بحث تختلف في محتوياتها بين فراغات البحث text search box، وعلامات البحث radios، وفراغات الترميز chek box...وغيرها. وعلى ذلك أصبح للزاحف حاجة في فهم نماذج البحث والتعامل معها تحليلاً وتكشيفاً، بغرض بناء علاقة وقواعد البيانات تجعل منها مصدرًا جديدًا يضاف إلى مجموعة مصادر الويب داخل مستودعات محركات البحث. وبسبب الأهمية الكبيرة التي يتسم بها عمل الزاحف في إخراج قواعد بيانات الويب من الظلمات إلى نور الويب المرئية، تتبنى الدراسة العناصر التالية في الحديث عن الزاحف. ويوضح الشكل (٣-٢) مدى التباين بين مهام الزاحف لصفحات الويب المرئية والزاحف لقواعد بيانات الويب غير المرئية[١].

---

(١) Ntoulas, Alexandros. Crawling and searching the hidden Web, supervising Gregory H. Leazer, UNIVERSITY OF CALIFORNIA, 2006, cited 208/2007, cited at http://www.dia.uniroma3.it/~vldbproc/017_129.pdf



الشكل (٣-٢) التباين بين مهام الزاحف في صفحات الويب وقواعد بيانات الويب غير المرئية

يوضح الشكل (٣-٢) اعتماد برنامج الزاحف في تتبع صفحات الويب على المسارات و الروابط، بينما يعتمد في بحث قواعد البينات على استخدام نماذج ثابتة داخله يتحكم من خلالها في ضم أو رفض واجهات بحث قواعد البيانات التي يصل إليها. ويؤدي برنامج الزاحف عمله أمام قواعد البيانات وفقاً للخطوات التالية(١):

١- استخدام مسارات واجهات البحث والوصول إليها.

٢- نسخ واجهة بحث قاعدة البيانات ثم تحليل محتواها من نماذج البحث وحقول البحث.

(١) Sriram Raghavan. Crawling the hiddenWeb, Stanford University,2001, cited 25/8/2007, cited at http://www.springerlink.com/index/jx4p4572l707k541.pdf

٣- قبول واجهة قاعدة بيانات الويب عند التأكد من نموذج البحث، أو إعادة تحليلها Form Analysis عند الاختلاف على موضوع نموذج البحث.

٤- يستخدم الزاحف استفسارات البحث لدراسة نموذج البحث والنتائج في واجهة قاعدة البيانات لتحديد القيم المسترجعة بها Response Analysis، وفي حالة وجود روابط يعمل برنامج الزاحف على الوصول إليها ودراستها Response Navigation:، ومن ثم قبول واجهة قاعدة البيانات أو رفضها مرة أخرى.

وتتشابه برامج الزاحف فيما بينها من حيث المهام وأسلوب العمل، ويكمن الاختلاف فقط في الشركات التابعة لها أو محركات البحث المضيفة. غير أنه يمكن القول إن الاختلاف بين برامج الزاحف أضحى أيضًا في بيئة العمل بين الويب المرئية وغير المرئية. وتوجد أمثلة كثيرة لبرامج زاحف الويب المرئية، غير أنه حتى الآن لم تخرج برامج أو نماذج تمثل زاحف الويب غير المرئية. ومن نماذج زاحف الويب المرئية(١):

1. Astra SiteManager2.0
2. COAST Web Master 7.0
3. MicrosoftSite Analyst 2.0
4. MicrosoftContent Analyzer 3.0Xenu1.2f
5. RBSE
6. WebCrawler
7. World Wide Web Worm
8. Google Crawler
9. CobWeb
10. Mercator

(١) Arroyo, Natalia. What is the What is the invisible web? invisible web? A crawler perspective, internet lab, cited at http://cybermetrics.wlv.ac.uk/AoIRASIST/Arroyo.Invisible_web_crawlers.pdf

أولاً: أنماط عمل الزاحف أو أشكال نسيج العنكبوت:

تتبنى برامج زحف الويب ثلاثة أساليب للتعامل مع مصادر الويب وصولاً وإضافة إلى مستودع محرك البحث. وتهتم برامج الزاحف في هذه الأساليب بعملية تصنيف مصادر الويب حتى تصل إلى مصادر الويب ذات البناء المخصص أو الموضوع المخصص أو المحتوى الشكلي المخصص، وما يشغلنا هنا هو إصباغ برنامج زاحف قواعد بيانات الويب غير المرئية للعمل بواحد من أساليب الزحف أو كلها حتى يكتشف مباشرة واجهات قواعد البيانات. وتتمثل هذه الأساليب في الزحف باستخدام مسار صفحة الويب أو الزحف باستخدام صفحة الويب ذاتها أو الزحف باستخدام نماذج واجهات الويب، وهو الأسلوب الذي يتوافق مع قواعد بيانات الويب غير المرئية. وهذه الأساليب كالتالي[1]:

الأسلوب الأول : الزحف والتصنيف بروابط الويب : Link Classifier

يعتمد الزاحف في هذا النمط على استخدام الخصائص التي ورد بها مسار أو رابط صفحة الويب. حيث تدل الكلمات الواردة به والعلامات المختلفة على طبيعة الصفحة الوارد منها أو الذاهب إليها هذا المسار. وتتمثل خصائص تحديد مسارات الويب وتصنيفها لدى الزاحف فيما يلي :

أولاً: مسار الويب ذاته وطبيعته URL حيث يختلف شكل ومفردات رابط الويب بين الصفحات النصية التقليدية وبين واجهات البحث أو صفحات نتائج البحث.

ثانيًا: وجهة رابط الويب anchor : حيث إن مسار صفحة البحث غالبًا ما يرتبط بصفحة النتائج، فطبيعة طرح الاستفسار هي تلقي الإجابة.

---

(١) Smith , Joan A., McCown, Frank and Nelson, Michael L. Observed web robot behavior on decaying web subsites, D-Lib Magazine, 2006, citd 1/9/2007, cited at http://www.dlib.org/dlib/february06/smith/02smith.html

ثالثًا: النص المحيط بالمسار text around the link وهو عبارة عن الكلمات المفتاحية المستخدمة في الاستفسار، بالإضافة إلى الكلمات الدالة على استراتيجية البحث المعرفة بواسطة صفحة البحث. وعلى ذلك يمكن القول إن برنامج الزاحف عندما يملك القدرة على تحليل طبيعة مسار أو رابط الويب فإنه عندئذ يملك كفاءة التعرف إلى واجهات بحث قواعد البيانات، ولا يحتاج الأمر سوى بعض التغيير في ملف تعليمات برنامج الزاحف.

الأسلوب الثاني : الزحف والتصنيف بصفحات الويب : Page Classifier

يضع برنامج الزاحف خصائص محددة للموضوع أو الموضوعات التي تنتمي إليها صفحة الويب، وذلك من خلال تحليل مكونات الصفحة ذاتها والنصوص أو تيجان الكلمات المفتاحية الواردة بهذه الصفحة. وهذا الأسلوب هو المستخدم من جانب محركات البحث عند القيام بعمليات التكشيف والتحليل لصفحات الويب حديثة الإضافة.

الأسلوب الثالث : الزحف والتصنيف بنماذج البحث : Form Classifier

تهدف برامج الزاحف في هذا الأسلوب إلى جمع نماذج واجهات قواعد البيانات، غير أن هناك الكثير من الواجهات التي تحتوي على نماذج بحث ولكنها ليست لقواعد البيانات، ومن أمثلة هذه النماذج صفحات الدخول لخدمات البريد الإلكتروني أو صفحات الانضمام لمجموعات المناقشة أو تلك الصفحات التي تحتوي على نماذج تسجيل لسلع تجارية ... وغيرها. وغالبًا ما تعترض هذه النماذج طريق برامج الزاحف عند العمل باستخدام نماذج البحث في تجميع قواعد بيانات الويب غير المرئية. ومن ثم فإن على برامج الزاحف التزود بالآليات التي تستطيع بها ترجيح وتحديد واجهات بحث قواعد البيانات دون غيرها من الواجهات الأخرى. ومن أهم هذه الآليات الاعتماد على شجرة القرار

decision tree. وتتبني هذه الآلية على تحديد مجموعة من الخصائص في الحكم على نماذج البحث، يستعين بها برنامج الزاحف واحدة تلو الأخرى حتى يصل إلى حكمه النهائي بتصنيف نماذج بحث قواعد البيانات معًا، وحجب نماذج البحث التي لا تنتمي لقواعد البيانات. وتتمثل هذه الخصائص في مجموعة من العناصر التي يتم حسابها للحكم على واجهة مصدر الويب، ويمثلها الجدول (٣-٣):

الجدول (٣-٣) عناصر التصنيف لنماذج البحث لدى برامج الزاحف

Decision tree elements	عناصر قرار التصنيف لبرنامج الزاحف
number of checkboxes	عدد حقول التصويب
number of radio tags	عدد حقول الاختيار من متعدد
number of file inputs	عدد حقول إدخال البيانات
number of submit tags	عدد تيجان إقرار المدخلات
number of image inputs	عدد حقول المواد غير النصية
number of buttons	عدد الأزرار في النموذج
number of resets	عدد مرات إعادة الإدخال
number of password tags	عدد تيجان كلمات المرور في النموذج
number of textboxes	عدد حقول إدخال البيانات النصية
number of items in selects	عدد وحدات الاختيار
sum of text sizes in textboxes	مجموع البيانات النصية في النموذج
presence of the string “search” within the form tag	حدد الحقول النشطة للبحث في النموذج

يستخدم برنامج الزاحف العناصر السابقة في الجدول (٣-٣) لتحديد قيمة نموذج البحث الذي يصل إليه الزاحف. ويمر برنامج الزاحف خلال عمله في بيئة

الويب بعدد من نماذج البحث التي تختلف في درجات احتوائها على العناصر، مما يدعم برنامج الزاحف في تحديد الواجهات ذات الصلاحية للقبول داخل محرك البحث. وقد دعم العمل بهذه العناصر بعض الآليات التي يبنى عليها برنامج الزاحف ومن هذه الآليات :

أولاً: الوصول الذاتي لواجهات البحث: Automated discovery of invisible databases

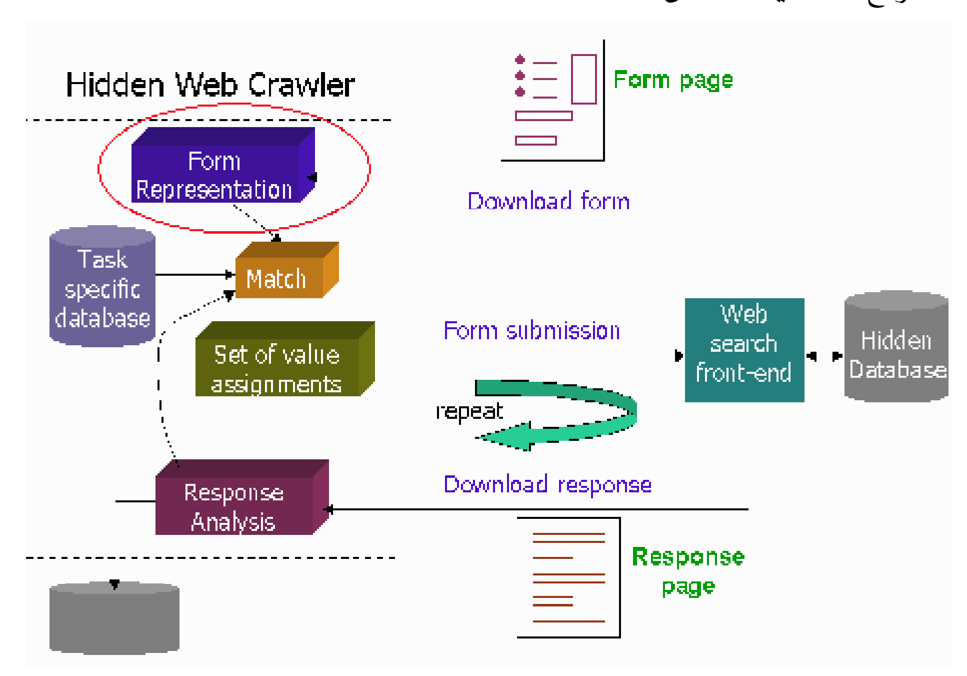
ظهرت هذه الفكرة على يد (Cope, 2003) لتمكين برامج الزاحف من أتمتة الوصول لواجهات بحث قواعد البيانات، وتسمى آلية العمل هنا بالسؤال السابق واللاحق لواجهة البحث pre-query and post query approach. ويستخدم برنامج الزاحف الاستفسار السابق لتحديد ملامح نموذج البحث في واجهة قواعد البيانات وتحديد جدوى إضافتها إلى مستودع محرك البحث، كما يوجه برنامج الزاحف استفسار البحث إلى واجهة قاعدة البيانات مباشرة ثم يقوم بدراسة خصائص شاشات عرض النتائج، وتسمى آلية العمل هنا بالاستفسار اللاحق. وتختلف طبيعة العبارات المستخدمة في استفسار البحث بحسب تقييم برنامج الزاحف لنموذج واجهة البحث؛ حيث يطلق على العبارات التي توافق الموضوع والعمل داخل قاعدة البيانات بالعبارات الإيجابية، وتسمى العبارات ذات النتائج الصفرية بالعبارات السلبية. ولقد أثبتت الاختبارات العملية لهذه الآلية أن واجهات البحث التي تنتمي بالفعل لقاعدة بيانات الويب تعطي النتائج ذات الخصائص التالية[1]:

- تتشابه واجهات عرض النتائج الصفرية تقريبًا في مساحة وحجم المعلومات المستخدمة والمعروضة.
- تستخدم واجهات عرض النتائج نموذجًا متشابهًا عند عرض وحدات النتائج؛ حيث تتكون واجهة العرض من شبكة أو شجرة من الحقول الدالة على وحدات النتائج المطابقة للبحث.

(١) invisible web. Closerlook, 2006, cited at 25/7/2007,cited at http://www.closerlooksearch.com/invisibleweb.aspx

ثانيًا: آليات بحث الزاحف لقواعد بيانات الويب:

يرى الباحث في النموذج التالي لبناء برنامج زاحف قواعد بيانات الويب غير المرئية مجموعة من الآليات التي يمكن تنفيذها داخل برنامج الزاحف حتى يتمكن من تحقيق أفضل أداء مع قواعد البيانات وصولاً وإضافة، اعتمادًا على خصائص عمل برنامج الزاحف السابق مناقشتها ومراحل العمل الموضحة في هذا النموذج كما في الشكل (٣-٣):



الشكل (٣-٣) نموذج عمل الزاحف على قواعد بيانات الويب

٢/٣/١/٢/٣ المعالجة والتفاعل مع قواعد بيانات الويب:

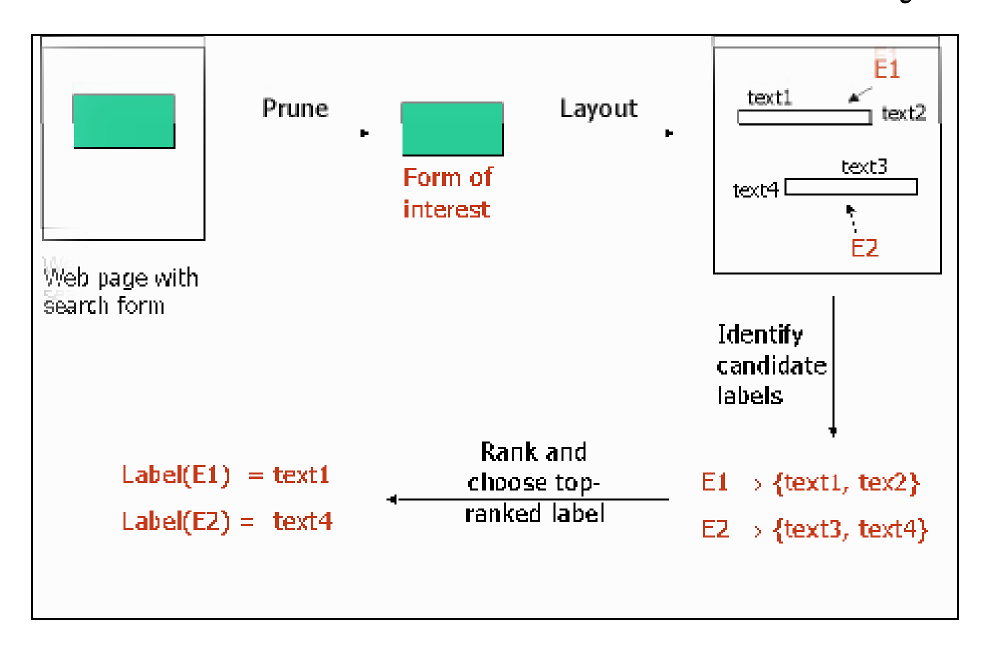
ينتهي عمل برنامج الزاحف مع قواعد بيانات الويب عند الوصول والإضافة إلى مستودع محرك البحث. وتبدأ المرحلة التالية لمعالجة استفسارات البحث وطرحها على قواعد البيانات ذات الصلة بالموضوع، ثم استقاء النتائج وعرضها على متصفح مستخدم الويب. ولكي تتم هذه العملية يحتاج محرك البحث إلى آلية

معالجة استفسارات ونتائج البحث من قواعد البيانات ذات التصميم المتباين وإليها. وتسمى هذه الآلية أو البرنامج بالوسيط Mediator ؛ حيث يقوم هذا البرنامج بدور الوساطة بين مستخدم محرك البحث وقواعد البيانات. وتتبنى برامج الوسيط وجهتين في التعامل ؛ تتعلق إحداهما بتحليل استفسار مستخدم الويب، أما الأخرى فتختص بالتعامل مع قواعد البيانات تحليلاً وتكشيفاً: حيث يأخذ الوسيط نسخة واجهة البحث القادمة من برنامج الزاحف لتحليل نموذج البحث الوارد بها والمكون من حقول البحث وعناوين هذه الحقول. ويخصص الوسيط درجة الاعتناء الأكبر بعناوين حقول البحث search labels، من خلال اقتباس عناوين حقول البحث يستطيع الوسيط تحديد الكلمات الدالة على موضوع قاعدة البيانات التي وردت في حقول البحث التي تعد بدورها مكونات التسجيلات الواردة داخل قاعدة البيانات. وتسمى آلية الاقتباس من تصميم واجهة البحث بـ LITE : Layout-Based Information Extraction. ويستخدم الوسيط آلية الاقتباس والتحليل هذه في موضعين من مهام عمل أداة البحث :

الأول : يأتي في تحليل نماذج بحث قواعد البيانات المضافة بواسطة الزاحف.

الثاني : في تحليل النتائج المستدعاة من قواعد البيانات الموافقة لموضوع البحث وترتيبها جميعًا في مجموعة واحدة لعرضها على متصفح مستخدم الويب. ويمثل الشكل (٣-٤) آلية الاقتباس والتحليل لعناوين حقول البحث في قواعد بيانات الويب غير المرئية[1].

(١) Raghavan, Sriram. Crawling the hidden web & other web-related research, stanford university, 2004, cited 22/8/2007, cited at http://www.acm.org/fcrc/PlenaryTalks/Garcia-Molina.pdf



الشكل (٣-٤) آلية الاقتباس والتحليل لعناوين حقول البحث في قواعد بيانات الويب غير المرئية

٣/٢/١/٣/٣ آليات محركات الويب لبحث قواعد البيانات:

تنتهي محركات بحث قواعد البيانات إلى مرحلة بحث واسترجاع محتوى الويب غير المرئية، وذلك بعد انتهاء مرحلتي الإضافة والمعالجة. وتبدأ محركات البحث عملها من حيث انتهت في التعامل مع محتوى قواعد بيانات الويب؛ حيث تبدأ مهامها من واجهة البحث واستقبال استفسارات البحث من مستخدمي الويب. وثمة عنصرين في عملية البحث يجب التركيز عليهما : العنصر الأول : نموذج واجهة بحث محرك قواعد البيانات، والعنصر الثاني: آليات البحث وتلقي النتائج.

العنصر الأول: نموذج واجهة البحث الرئيسية لمحرك البحث:

تتوسط محركات البحث الاتصال بين مستخدم الويب ومحتوى قواعد البيانات، وعلى ذلك فتحتاج محركات البحث إلى إدراك تام لمخططات واجهات ونماذج بحث قواعد البيانات databases search schemas ذات التصميم المختلف،

ومن ثم صياغة نموذج عام للبحث يوضع أمام مستخدم الويب ويصلح في الوقت ذاته إلى التعامل مع مختلف نماذج بحث قواعد البيانات.

وحمل بعض الباحثين على عاتقهم محاولة إدراك وتحليل نماذج بحث قواعد البيانات حتى يتسنّى لنظم استرجاع قواعد البيانات طرح استفسارات البحث بما يتوافق وآليات البحث في واجهات قواعد البيانات. ومن هؤلاء Hai He وآخرون اللذين عملوا على تصميم نموذج سمي بـ (WISE-iExtractor) ، ويتكون هذا النموذج من شقين هما ؛

الأول : Attribute Extraction .

والثاني: Attribute Analysis.

حيث يقوم الشق الأول باقتباس عناصر البحث المترابطة في نماذج واجهات قواعد البيانات ؛ ومثال لهذا الترابط ما يوجد عيه حقل تاريخ النشر من إعطاء فترة زمنية بين قيمتين موزعتين على عنصري بحث تحت عنوان "قبل" و"بعد". ثم يقوم الشق الثاني من WISE-iExtractor بتحليل خصائص وملامح هذه العناصر لاكتشاف الروابط والدلالات المنطقية المؤثرة في تشكيل واجهة قاعدة البيانات كلها. ويرى الباحث أن استخدام هذا النموذج داخل محركات بحث قواعد البيانات يمكن محركات البحث من فهم خصائص عناوين عناصر البحث labels، بجانب تحديد خصائص قيم البيانات التي توضع بها data values. وذلك لبناء واجهة بحث عامة تتوافق مع خصائص وقيم عناصر البحث في قواعد البيانات المختلفة[1].

---

(١) Hai He and Meng, Weiyi. Constructing interface schemas for search interfaces of web databases, 2005, cited 208/2007, cited at www.cs.binghamton.edu/~meng/pub.d/He_p136.pdf

أنماط العلاقات بين عناصر البحث في واجهات قواعد البيانات: Logic relationship of attributes

ترسم عناصر البحث داخل واجهات قواعد البيانات شكلاً أو مخططًا متكاملاً، ترتبط فيه عناصر البحث بأنماط ربط منطقية مختلفة تتمثل فيما:

العناصر المتواصلة: Conjunctive

حيث ترتبط هذه العناصر بمعامل الربط البولياني AND ، مما يعني أنها جميعًا واجبة الإكمال والإدخال، ولا يمكن الاستغناء بأحد منها عن الآخر.

العناصر المنفصلة: Disjunctive

وترتبط عناصر البحث هنا بمعامل الربط OR ، حيث يكفي في عملية البحث استخدام أي منها، أو أن واحدًا من هذه العناصر بمفرده يكفي لإجراء عملية البحث.

العناصر المحددة : Exclusive

وتتواجد هذه العناصر جميعها في واجهة البحث كاملة غير أن عملية البحث تقتصر على اختيار أحدها فقط، وتتمثل هذه العناصر في radio buttons.

العناصر المهجنة المتداخلة : Hybrid

ويتمثل هذا الترابط في الجمع بين أغلب أنواع الترابط السابقة أو كلها.

ويمكن القول إن عمل نموذج WISE-iExtractor يمر بمرحلتين، تتكون كل منهما من مجموعة من الخطوات التي يمكن تلخيصها في النقاط التالية[١]:

☒ **المرحلة الأولى : اقتباس خصائص البحث** Attribute Extraction

❖ تعيين معنى أو شارات واجهة البحث Interface Expression

---

(١) Hai He and Meng, Weiyi. Ibid.

- ❖ اقتباس عناصر وعناوين حقول البحث المنفردة Extracting Individual Labels and Elements
- ❖ تحديد مسميات حقول البحث المقتبسة Identifying the Names of Exclusive Attributes
- ❖ تقسيم عناوين وعناصر البحث المقتبسة إلى مجموعات Grouping Labels and Elements

حيث يعمل نموذج WISE-iExtractor على تنظيم عناوين البحث المقتبسة في مجموعات، تتشابه كل مجموعة فيما بينها على أساس وظيفة البحث ؛ مثل حقول البحث عن المؤلف في الاسم الأول والأخير واللقب. وتتكون هذه المجموعات تبعًا لخمسة مقاييس أساسية في تنظيم عناصر البحث في مجموعات مختلفة، وهي:

١- الحقول النهائية لعناصر البحث ؛ حيث دائمًا ما يعبر كل عنصر بحث عن حقل من الحقول المكونة لتسجيلات قاعدة البيانات.

٢- مدى التشابه في النص بين مسمى عنصر البحث وبين محتوى الكلمات البحثية الداخلة فيه.

٣- مدى القرب أو البعد بين مسمى أو عنوان عنصر البحث وبين فراغ البحث المتاح لإدخال الكلمات البحثية، وهو ما عكس توزيع العناصر البحثية وأهمية عنونتها.

٤- استقامة عنوان وعنصر البحث في صف واحد أم أنها توزع على صفين في واجهة البحث.

٥- أولوية ترتيب عناصر البحث ؛ حيث يعني تقديم عنصر على آخر قدرًا أكبر من الأهمية في إجراء بحث قاعدة البيانات.

☒ **المرحلة الثانية : تحليل خصائص وعناصر البحث** Attribute Analysis

❖ تمييز عناصر البحث وخصائصه فيما بينها Differentiating Elements

❖ التحديد والتعريف بالروابط المنطقية بين عناصر البحث وخصائصه Identifying the Relationships and Semantics of Domain Elements

❖ اشتقاق الدلالة أو المعنى العام لخصائص البحث وعناصره Deriving Meta-information of Attributes : وذلك بالاعتماد على أربعة عوامل أساسية كالتالي:

١-نوع مجال البحث المتمثل في طبيعة معاملات الربط المستخدمة في صياغة الكلمات المفتاحية في البحث.

٢-نوع البيان المستخدم، سواء أكان نصًا أم عملةً أم رقمًا أم تاريخًا ووقتًا ... وغيرها.

٣-القيمة المحددة المرجحة default التي تستخدمها واجهة البحث في بعض الحقول.

٤-الوحدة القياسية المستخدمة لبعض بيانات الحقول، مثل وحدات المتر للأطوال، والكيلوجرام للأوزان، والجنية للعملة...وغيرها.

❖ تحديد الروابط الدلالية والمنطقة بين عناصر وخصائص البحث Identifying the Logic Relationships

العنصر الثاني : آليات البحث وتلقي النتائج في محرك البحث:

يعمل محرك بحث قواعد البيانات هنا بالشكل الذي يوافق عمل محركات البحث المتعددة Meta search engines، حيث يتلقى محرك قواعد البيانات استفسارًا ثم يرسله إلى قاعدة أو قواعد البيانات الملائمة موضوعيًا، ثم يتلقى النتائج ويعرضها أمام مستخدمي الويب، تمامًا مثلما تفعل محركات البحث

المتعددة مع محركات بحث الويب. ومن ثم فإن آلية العمل هي ذاتها المستخدمة في محركات البحث المتعددة. فضلاً عن أن منهجنا في هذه الدراسة هو استخدام آليات عمل محركات البحث المتعددة. ويرى الباحث أن هذا الجزء من الدراسة سوف يتم تناوله في نهاية هذا الفصل عند الحديث عن محركات البحث المتعددة لتطابقهما في المعنى والهدف والمحتوى.

لقد دفعت قواعد بيانات الويب غير المرئية عددًا من الباحثين إلى التفكير في إيجاد نظام متكامل لقواعد البيانات، يمكن مستخدمي الويب من بحث مختلف قواعد بيانات الموضوع أو التخصص الواحد باستخدام واجهة بحث واحدة. وقد كان نظام EasyQuerier من بين الحلول المقترحة للنظم المتكاملة لبحث قواعد البيانات. ويعمل هذا النظام على استخدام واجهة بحث واحدة في بحث أكثر من قاعدة بيانات ذات التخصص الواحد واسترجاع نتائج البحث منها جميعًا في وقت واحد وفي نافذة واحدة. وقد خرج هذا النظام موجهًا لعلاج مشكلة قواعد البيانات محددًا ركائز المشكلة فيما يلي[(١)] :

١- إن مستخدم الويب يحتاج أولاً إلى تحديد المجال الموضوعي المحدد موضع البحث بدقة، حتى يتمكن من التوجه إلى واجهة البحث الصحيحة، وهذا ما يزداد صعوبة يومًا بعد آخر في ظل نمو موضوعات المعرفة البشرية.

٢- إن واجهة البحث الرئيسة للنظام المتكامل سوف تكون على درجة من التعقيد التي قد لا تتناسب مع بعض مستخدمي الويب، وذلك يرجع إلى احتواء هذه الواجهة على خصائص بحث متداخلة معقدة، إضافة إلى قيم بحثية سابقة التحديد default values.

(١) Li, Xian and Meng, Weiyi. EasyQuerier: A Keyword based interface for web database integration system, Renmin University of China, 2007, cited 10/9/2007, cited at http://www.springerlink.com/index/jx4p4572l707k541.pdf

٣- على مستخدم الويب أن يجري عملية البحث أكثر من مرة، كلما احتاج إلى تغيير القيم البحثية الداخلة إلى واجهة البحث ؛ حيث لا تقبل واجهة البحث إدخال أكثر من قيمة في كل عنصر بحث.

ويتحدد عمل نظام EasyQuerier في التعرف إلى قواعد البيانات ذات التخصص الواحد، ثم تطبيق أساليب الاقتباس والتحليل لنماذج واجهات بحث قواعد البيانات، وتأتي المرحلة اللاحقة متمثلة في تصميم واجهة واحدة مكونة من حقول بحث ثابتة يستطيع مستخدم الويب من خلالها بحث مختلف قواعد البيانات ذات التخصص الواحد، ثم إخراج النتائج على الواجهة الرئيسية للنظام المتكامل.

٢/٢/٣ أدلة بحث قواعد بيانات الويب غير المرئية:

١/٢/٢/٣ مفهوم دليل البحث:

يعرف قاموس مصلحات المكتبات والمعلومات على الخط المباشر ODLIS دليل البحث web directory على أنه " إحدى خدمات الخط المباشر التي تكشف مواقع ومصادر الويب المتاحة على الإنترنت، وهي تتيح الوصول لمصادر الويب من خلال اتباع نظام تصنيف هرمي للموضوعات، ومن أهم أدلة الويب دليل Yahoo."
وطبقًا لذلك فإننا نعرف دليل الويب غير المرئية على انه إحدى أدوات بحث مصادر الويب غير المرئية مثل قواعد البيانات وغيرها، ويعمل على استدعاء هذه المصادر باستخدام إحدى قوائم رؤوس الموضوعات أو خطط التصنيف المرتبة منطقيًا.
وتعتمد أدلة البحث على العنصر البشري في معالجة مصادر الويب وتنظيمها، خلاف محركات البحث التي تعتمد كلية على برامج فرعية مثل الزاحف والمكشف واليات البحث[١].

---

(١) Reitz, Joan M. directory, ODLIS — Online Dictionary for library and information science, libraries unlimited, 2006, cited 20/8/2007, cited at http://lu.com/odlis/odlis_d.cfm#directory

وتتسم أدلة البحث بملامح خاصة تكمن فيما يلي[1]:

- ☒ **الحفاظ على السياق** : حيث يتم تكشيف مصادر الويب تبعًا للمحتوى الفعلي للمصدر، مما يجنب حدوث أخطاء البحث المختلفة. وعلى سبيل المثال عند اختيار مصطلح "أحمد زويل" كموضوع للبحث، فسوف يتجنب دليل البحث استدعاء مؤلفات أحمد زويل، وإنما سيكون البحث عن زويل كموضوع.
- ☒ **جودة الأداء** : ويرجع ذلك إلى استخدام العنصر البشري في كل مراحل عمل دليل البحث، بما يتفادى معه أخطاء الحاسب الآلي في معالجة محتوى مصادر الويب.
- ☒ **إمكانات التصفح** : صممت أدلة البحث بما يوفر لمستخدم الويب تصفح الموضوعات والاختيار فيما بينها بما يلائم درجة تخصص موضوع البحث.
- ☒ **درجة الحداثة** : يسبب اعتماد أدلة البحث على العنصر البشري بعض التأخر في متابعة كل حديث على الويب.
- ☒ **حجم قاعدة البيانات** : تتسم قاعدة بيانات دليل البحث بمحتوى أصغر حجمًا عنها في محرك البحث.
- ☒ **التوضيحات والحواشي** : تضيف أدلة البحث بعض حواشي الحكم على محتوى مصدر الويب بجانب الكلمات الدالة على هذا المحتوى.

٢/٢/٢/٣ أدلة البحث وقواعد البيانات:

يرتبط مصطلح دليل الويب دائمًا بالتقسيم والتنظيم الموضوعي لمصادر الويب أسفل قوائم مصطلحات أو رؤوس موضوعات، تتدرج هذه القوائم من العام إلى

---

(١) searching internet resources.cited 20/9/2007, cited at http://qprober.cs.columbia.edu/publications/vldb2002.pdf

الخاص إلى الأكثر تخصصًا، ويرتبط كل مصطلح لها بما يتصل به من مصادر الويب. وتعتمد أدلة الويب على العنصر البشري في كل مراحل معالجة مصادر الويب بدءاً من الإضافة، وانتهاءً بتصفح قوائم رؤوس الموضوعات المعدة من قبل الإنسان. ويرتبط دليل الويب دائمًا باستخدام أسلوب التصفح في بحث مصادر الويب واسترجاعها، مثلما يرتبط محرك البحث باستخدام أسلوب البحث الحر بالكلمات المفتاحية. ويمكن القول إن أدلة البحث هي القاسم المشترك لتنظيم قواعد بيانات الويب غير المرئية مع محركات البحث، من خلال تحليل محتوى قواعد البيانات والتعبير عنها في شكل مصطلحات تلائم المجال الموضوعي لكل قاعدة أو مجموعة قواعد.

وتكمن مشكلة قواعد البيانات مع الأدلة في غياب وسائل الربط أو الوسيط بين قوائم رؤوس الموضوعات، التي تعد العمود الفقري لعمل الأدلة، وبين التسجيلات أو الوحدات المكونة للمحتوى الداخلي لقواعد البيانات. فبخلاف محركات البحث لا يمكن لأدلة الويب نقل المصطلحات البحثية أو استراتيجية البحث إلى واجهات بحث قواعد البيانات، ومن ثم يكون مستوى البحث والاسترجاع متوقفاً عند استرجع قواعد البيانات كمصادر، دون القدرة على أن تمد أدلة البحث يدها إلى عمق المحتوى الداخلي من تسجيلات وجداول داخل قاعدة البيانات. وعلى ذلك فمن المقبول لدينا أن نرى محرك بحث قواعد بيانات بمفرده دون الحاجة إلى دليل، ولكن من غير المقبول أن نرى دليل بحث قواعد بيانات بمفرده دون استخدام أسلوب البحث الحر بجانبه.

لقد دعت هذه المشكلة الباحث إلى تبني فكرة حتمية الجمع بين آليات عمل محرك البحث ودليل البحث في نظام واحد لاسترجاع قواعد بيانات الويب غير المرئية. ولذلك فإن هذا الجزء من الدراسة يتكامل مع جزئية الحديث عن آليات عمل محرك بحث قواعد البيانات. ومن ثم فإن أدلة البحث هي الشق الثاني في

تكوين نظام استرجاع قواعد البيانات وليس كنظام منفرد. بالإضافة إلى ذلك فإن أدلة بحث الويب غير المرئية يمكن أن تتسم بالقصور في الجوانب التالية:

- تستخدم أدلة البحث العنصر البشري في عمليات المعالجة لمصادر وقواعد بيانات الويب، مما يجعل درجة الأداء متوقفة على كفاءة ومستوى العنصر البشري القائم على المعالجة، إلى جانب عدم تمكن العنصر البشري بقدرته المحدودة من مسايرة النمو السريع لمصادر الويب غير المرئية.
- تحيل أدلة الويب المرئية مستخدميها إلى قائمة بمواقع الويب التي تخدم الموضوع المخصص، وهو ما لا يتوافق مع بحث مواقع الويب غير المرئية التي تعتمد في معظمها على واجهات بحث لاستدعاء محتواها.

إن ثمة تطور قد طرأ على أدلة الويب أداءً ومفهومًا يمكنها من أن تصل إلى إمكانية التعامل مع الويب غير المرئية ؛ أما التطور في الأداء فيتمثل في إمكانية استخدام العنصر الآلي لتحليل محتوى الويب غير المرئية (قواعد البيانات) بجانب الاعتماد على العنصر البشري الذي تختص به أدلة البحث عن محركات البحث.

وأما التطور في المفاهيم فيتمثل في ارتباط دليل الويب وعملية التقسيم الموضوعي التي يبنى عليها عمله بمجموعة من المفاهيم التي ظهرت في الأدبيات الأجنبية. ويعد مصطلح data clustering التقسيم الموضوعي للبيانات، أفضل أمثلة تطور المفاهيم. ويعني هذا المفهوم كما عرفه Pavel Berkhin بأنه " تقسيم البيانات إلى مجموعات بحسب التشابه في الموضوع، وتعرض مجموعات البيانات من خلال مصطلحات قد يصاحبها إيضاحات في بعض الأحيان. وغالبًا ما يكون هذا العرض في شكل بناء هرمي يجمع بين هذه المصطلحات الدالة على الموضوعات، ويطلق على كل منها عقدة node." وقد ارتبط مفهوم data clustering بمفهوم آخر هو dat mining الذي يمكن ترجمته إلى البحث والتنقيب عن المعلومات وذلك في مرحلة البحث عنها، أو ترجمته إلى تخزين

وتنظيم المعلومات في مرحلة المعالجة. وقد حدد Pavel Berkhin أيضًا في دراسته عن التقسيم الموضوعي للبيانات clustering بعض إيجابياته وسلبياته كما يلي[1]:

فإيجابياته هي:

١- تمثيل المعلومات والبيانات في شكل تنظيم مرن يسهل التحرك خلاله.

٢- سهولة التمييز بين المجموعات المتصلة موضوعيًا وغير المتصلة.

٣- التتابع والترابط بين مستويات الموضوع المختلفة.

وسلبياته هي:

١- عدم وجود المقاييس المقننة لبناء التقسيمات الموضوعية ؛ حيث تستخدم عادة وجهة نظر القائمين عليها.

٢- تتسم معظم أشكال البناء الهرمي بالتباطؤ في تحديثها ، لتباعد فترات تقييمها وزيارتها.

٣- عدم وجود الترابط والعلاقات المنهجية بين الموضوعات.

٣/٢/٢/٣ تكشيف قواعد البيانات داخل أدلة البحث:

يؤكد واقع الويب أن تكشيف مصادر الويب المرئية داخل أدلة البحث يختلف عنه في مصادر الويب غير المرئية. بل إن الأمر يتعدى ذلك لتختلف مصادر الويب غير المرئية فيما بينها في عملية التكشيف داخل أدلة بحث الويب؛ فتحتاج قواعد البيانات إلى خصائص تكشيف تختلف عن مواقع الويب التفاعلية مثلاً ؛ لأن واجهة موقع الويب قد تعطي بدورها البيانات المساعدة في تكشيف المصدر ،

(١) Berkhin, Pavel. Survey of clustering data mining techniques. Accrue Software, Inc., 2003, cited 20/9/2007, cited at
http://www.ee.ucr.edu/~barth/EE242/clustering_survey.pdf
searching internet resources

خلاف ما توجد عليه واجهة قاعدة البيانات التي لا تعطي معلومات سوى إرسال استفسارات البحث لها واستقبال النتائج منها.

ومن أبرز النظم المستخدمة في تحليل محتوى الويب غير المرئية وتكشيفها، نظام The QProber system الذي قدمته Columbia University، ويعمل من خلال تحديد ملامح واجهات مواقع الويب غير المرئية اعتمادًا على نظام تصنيف محدد سلفاً، ثم إرسال استفسارات البحث لهذه المواقع، وتحليل النتائج المستدعاة لإخراج الكلمات الدالة منها. غير أن هذا النظام قد يعمل بكفاءة مع أدلة بحث قواعد البيانات في مرحلة التكشيف والمعالجة، أما في مرحلة البحث والاستدعاء فتظل أدلة البحث قاصرة عن الوصول المباشر إلى محتوى قواعد البيانات واسترجاعه أمام مستخدمي الويب[١].

وتحتاج قواعد بيانات الويب غير المرئية إلى معالجة وتكشيف تتسم بالمراحل التالية[٢]:

١- إرسال الاستفسار الأولي لواجهة قاعدة البيانات حاملاً أعم المصطلحات المستخدمة في قوائم دليل البحث.

٢- تحليل أعداد النتائج المسترجعة لكل مجموعة مصطلحات عامة أو خاصة.

٣- حساب النتائج ذات الأعداد الكبيرة التي تغطي كل مصطلح على حدة.

٤- وضع وتسكين قاعدة البيانات تحت أكثر المصطلحات استرجاعًا للنتائج.

٥- تكرار تسكين قاعدة أو قواعد البيانات تحت كل مصطلح من المصطلحات واسعة الاستدعاء.

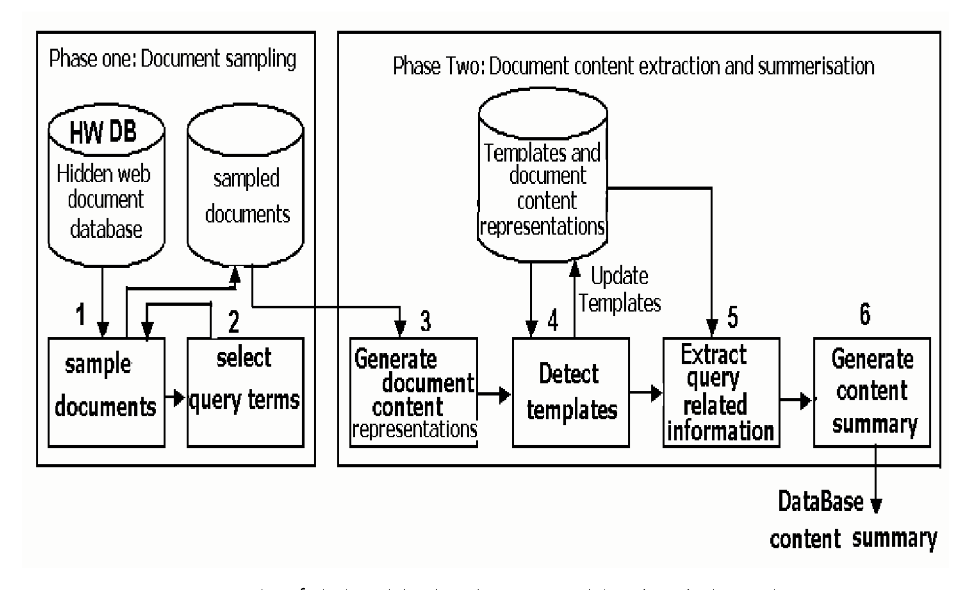
(١) GRAVANO, LUIS and IPEIROTIS, PANAGIOTIS G. Qprober.op.cit.

(٢) Ipeirotis, Panagiotis G. and Gravano, Luis. Distributed search over the hidden web: hierarchical database sampling and selection, Columbia University, 2002, cited 10/8/2007, cited at http://qprober.cs.columbia.edu/publications/vldb2002.pdf

٦-استخدام مختلف المصطلحات التي تنتمي أو الدرجة نفسها في القوائم الموضوعية لدليل البحث.

٧-تؤكد النتائج المسترجعة وأعدادها على أكثر المصطلحات تطابقًا مع قاعدة البيانات، ومن ثم يتم الربط بين قاعدة البيانات وبين المصطلح أو المصطلحات الدالة على المحتوى.

ويوضح الشكل (٣-٥) مراحل العمل السابقة لتكشيف وتحليل محتوى قواعد البيانات داخل أدلة البحث.



الشكل (٣-٥) تحليل محتوى قواعد البيانات داخل أدلة البحث.

يوضح الشكل (٣-٥) مراحل التكشيف والتحليل الآلي لقواعد البيانات داخل أدلة البحث، حيث تستطيع البرامج الآلية الوصول إلى النقاط التي تكبد العنصر البشري وقتًا أطول وجهدًا اكبر. فضلاً عن أن العنصر البشري سيكون له الصبغة النهائية لضبط وتقنين الكلمات الدالة الخارجة من هذه المعالجة الآلية

لقاعدة البيانات. وتوضح الأرقام في الشكل السابق الخطوات التي تمر بها عملية التكشيف موزعة على مرحلتين أساسيتين كما يلي(١):

المرحلة الأولى : معاينة (تسجيلات) قاعدة البيانات document sampling:

١- الحصول على عينة من وحدات أو وثائق محتوى قاعدة البيانات.

٢- تحديد أهم المصطلحات التي ستستخدم لطرح استفسارات البحث واسترجاع المحتوى الداخلي.

المرحلة الثانية : اقتباس الكلمات الدالة extarct content summary:

٣- تجميع مختلف المصطلحات ذات الدلالة الأكبر على محتوى قاعدة البيانات.

٤- تحديث وضبط المصطلحات الدالة ذات الارتباط الموضوعي وتقنينها للبحث بها مرة أخرى.

٥- اقتباس المصطلحات الدالة ذات العلاقة الموضوعية المباشرة بموضوعات استفسارات البحث.

٦- ضبط وتجميع مختلف المصطلحات الموضوعية وربطها بقاعدة أو قواعد البيانات ذات العلاقة، وحفظها داخل مستودع دليل الويب.

٤/٢/٢/٣ تقسيم محتوى قواعد البيانات متعددة الوسائط:

تتطابق قواعد البيانات متعددة الوسائط مع قواعد البيانات النصية من حيث بنية المحتوى في استخدام التسجيلات الببليوجرافية. غير أنها تتميز بخصائص التعامل مع الوسائط المتعددة من ملفات الصورة والصوت، حتى تتمكن من تحليل

(١) Hedley,Yih-Ling and Younas, Muhammad. Sampling, information extraction and summarisation of hidden web databases, School of Mathematical and Information Sciences, Coventry University, 2006, cited 15/9/2007, cited at http://portal.acm.org/citation.cfm?id=1225816.1225818

ومعالجة وسيط حمل المعلومات من صوت وصورة. وعلى ذلك فإن ثمة أسلوبين تستطيع بهما أدلة البحث التعامل مع قواعد البيانات غير النصية، هما: أسلوب **التحليل المبني على النص**، وهو ما تم تناوله في أدلة البحث. وأسلوب **التحليل المبني على المحتوى** CBR. ويقتضي هذا الأسلوب تزويد أدلة بحث قواعد البيانات بمجموعة من آليات التعامل مع الوسائط غير النصية وتحليل البنية واللون والشكل Texture, Color and Shape. إلى جانب استخدام التحليل النصي للحقول المكونة لتسجيلات هذه الوحدات داخل قواعد البيانات.

٣/٢/٣ بوابات قواعد البيانات:

يوصف مصطلح البوابة protal or gatway بأنه القديم الحديث في الويب؛ حيث يعني هذا المصطلح مواقع الويب التي توفر خدمات عدة لمستخدمي الويب، من بينها البحث والاسترجاع لمصادر الويب، ويمكن أن ينسحب مفهوم البوابة على النظام المتكامل لبحث واسترجاع قواعد بيانات الويب غير المرئية الذي تسعى إلى تحقيقه هذه الدراسة، لأن النظام المقترح يحقق الشمولية في بحث قواعد بيانات الويب باستخدام البحث الحر في محركات البحث والبحث بالتقسيم في أدلة البحث، بالإضافة إلى خدمات أخرى تتعلق بالويب غير المرئية. وقد عمد الباحث إلى تناولها حتى لا ينتهي العمل إليها ولم يرد ذكرها، كما أنها أهم أدوات البحث على الويب مشاركة مع محركات وأدلة بحث الويب. وقد عرف عبد الرحمن فراج البوابة على أنها" عبارة عن قاعدة بيانات تشتمل على تسجيلات مفصلة لما وراء البيانات detailed metadata records ، تقوم بوصف مصادر الإنترنت وتوفير الرابطة الفائقة Links لهذه المصادر. والمستفيد هنا يمكنه الاختيار فيما بين البحث في قاعدة البيانات بواسطة الكلمات الدالة، أو تصفح المصادر تحت رؤوس الموضوعات." إلى جانب تعريفها بأن " البوابة (portal) : هي موقع ويب يُشَكِّل نقطةَ البداية للاتصال بمواقع الويب الأخرى، وقد جاء اسم البوابة من وظيفتها كباب مفتوح يُطل المرء منه على عالم المعلومات والفعاليات

الأخرى التي تُوفرها الإنترنت. وتتميز البوابة عن مواقع الويب بدرجة عالية جدًا من التنظيم" [1] [2].

وتقسم البوابات إلى أنواع مختلفة اعتمادًا على طبيعة المجال أو قطاع المستخدمين الذي تستهدفه:

١- البوابات العمودية (vertical or niche) هي بوابات تتناول شريحة محددة من المستفيدين، إذ تُركِّزُ المعلومات الموجودة فيها على هدف أو اتجاه موضوعي آخر.

٢- البوابات القطاعية (industry portal) التي تُدعى أيضاً بوابات الأعمال التي تخدم قطاع الأعمال (Business-to-business- B2B) وهي تشبه بوابات معلومات العملEIP .

٣- البواباتُ الأفقية (horizontal portal) التي تُقدم مجموعةً واسعة ومتنوعة من المواد والموضوعات ذات الطابع العام، وتنطوي هذه الفئة على الكثير من البوابات التي نذكر منها : ياهو (Yahoo!).

٣/٣ محركات البحث المتعددة لقواعد البيانات:

تتبنى الدراسة فكرة تطبيق البحث المتعدد في بحث واسترجاع الويب غير المرئية. وذلك كما يلي:

١/٣/٣ مفهوم محركات البحث المتعددة:

يعرف قاموس المكتبات والمعلومات على الخط المباشر ODLIS البحث المتعدد على أنه " بحث للمعلومات باستخدام برامج صممت لبحث واسترجاع المعلومات

(١) عبد الرحمن فراج. البوابات ودورها في الإفادة من المعلومات المتاحة على الإنترنت، مجلة المعلوماتية، ٢٠٠٠م، تاريخ الاستشهاد ٢٠٠٧/٨/٢٥، مسار
http://informatics.gov.sa/modules.php?name=Sections&op=viewarticle&artid=47

(٢) البوابات(Portals) . مشروع سمو الشيخ محمد بن راشد آل مكتوم لتعليم تكنولوجيا المعلومات، ٢٠٠٥م، تاريخ الاستشهاد ٢٠٠٧/٧/٢٠، مسار
http://www.itep.ae/arabic/EducationalCenter/Articles/portal_01.asp

من مجموعة محركات بحث مختلفة ثم دمج وإعادة ترتيب النتائج المسترجعة. ومن أهم أمثلة محركات البحث المتعددة على الويب Dogpile, Monster Crawler, and ProFusion". أيضًا عرفت شركة Blue Angel Technologies عملية البحث المتعدد على أنها " عملية بحث المعلومات اعتمادًا على طرح سؤال بحث واحد لبحث مجموعة من نظم المعلومات المختلفة في وقت واحد. ومن الممكن أن تشمل هذه النظم فهارس المكتبات وقواعد البيانات التجارية والمجموعات الرقمية والمجموعات الخاصة". وقد ركز أحد التعريفات الأخرى على محرك البحث ذاته دون عملية البحث كما سبق في المفهومين السابقين ؛ حيث عُرف محرك البحث المتعدد على أنه " أحد أنواع أدوات البحث التي تمكن مستخدم الويب من بحث مجموعة من اثنين أو أكثر من محركات وأدلة البحث في وقت واحد. وهذه الأداة لا تملك كشافًا لمصادر المعلومات خاصًا بها ، وإنما صُممت لتكون على قمة أدوات بحث الويب الأخرى". ويرى الباحث أن هذا التعريف الأخير يعكس بدقة ماهية محرك البحث المتعدد(١)(٢)(٣).

خرجت محركات البحث المتعددة للعمل مع قاعدة بيانات محركات البحث ، ويتمثل العمل في استقبال استفسار البحث من مستخدمي الويب ثم إرساله إلى محركات البحث لاستدعاء كل ما تحمله محركات البحث المختلفة عن استفسار المستخدم. ومن ثم يعطي محرك البحث المتعدد لمستخدم الويب إمكانية البحث في جميع محركات البحث مرة واحدة. أي أنه يعمل غالبًا كواجهة بحث رئيسية وشاشة لعرض النتائج المسترجعة أمام مستخدم الويب.

---

(١) Reitz, Joan M. metasearch, ODLIS — Online dictionary for library and information science, Libraries Unlimited, 2006, cited 20/9/2007, cited at http://lu.com/odlis/odlis_m.cfm# metasearch

(٢) metasearch.Blue angel technologies, Inc.2005, cited 25/9/2007, cited at http://www.BlueAngelTech.com

(٣) metasearch engines. Clixmarketing, cited 10/10/2007, cited at www.clixmarketing.com/definitions.htm

٢/٣/٢ محركات البحث المتعددة وقواعد بيانات الويب:

يتساوى محرك بحث قواعد البيانات في الويب غير المرئية مع محرك البحث المتعدد في الويب المرئية. حيث يبحث محرك البحث المتعدد في قاعدة بيانات محركات البحث، ويبحث محرك بحث الويب غير المرئية في قواعد بيانات الويب. واعتمادًا على ذلك فإن خصائص العمل في محركات البحث المتعددة يمكن أن تنسحب على محرك بحث قواعد البيانات الذي تهدف إليه هذه الدراسة. وكان لزامًا على الباحث دراسة هذه الخصائص لإصباغ محرك بحث قواعد البيانات بها، خاصةً، لتشابه حيثيات العمل بين محرك البحث المتعدد ومحرك قواعد البيانات.

يستخدم الباحث آليات ومكونات عمل محركات البحث المتعددة في هذا الجزء لتطبيقها على بحث واسترجاع محتوى قواعد بيانات الويب. وثمة عوامل دعمت هذا الاتجاه تمثلت في الأسباب الآتية(١):

١- تتسم الويب عامة وقواعد البيانات خاصة بتغطية واسعة وشاملة لمختلف الموضوعات، مما أوجب الحاجة لأدوات بحث تستطيع تطبيق مبدأ الشمول والتغطية في عمليات البحث. وتعد آلية البحث المتعددة هي أفضل آليات التعامل مع بحث واستدعاء محتوى قواعد بيانات الويب.

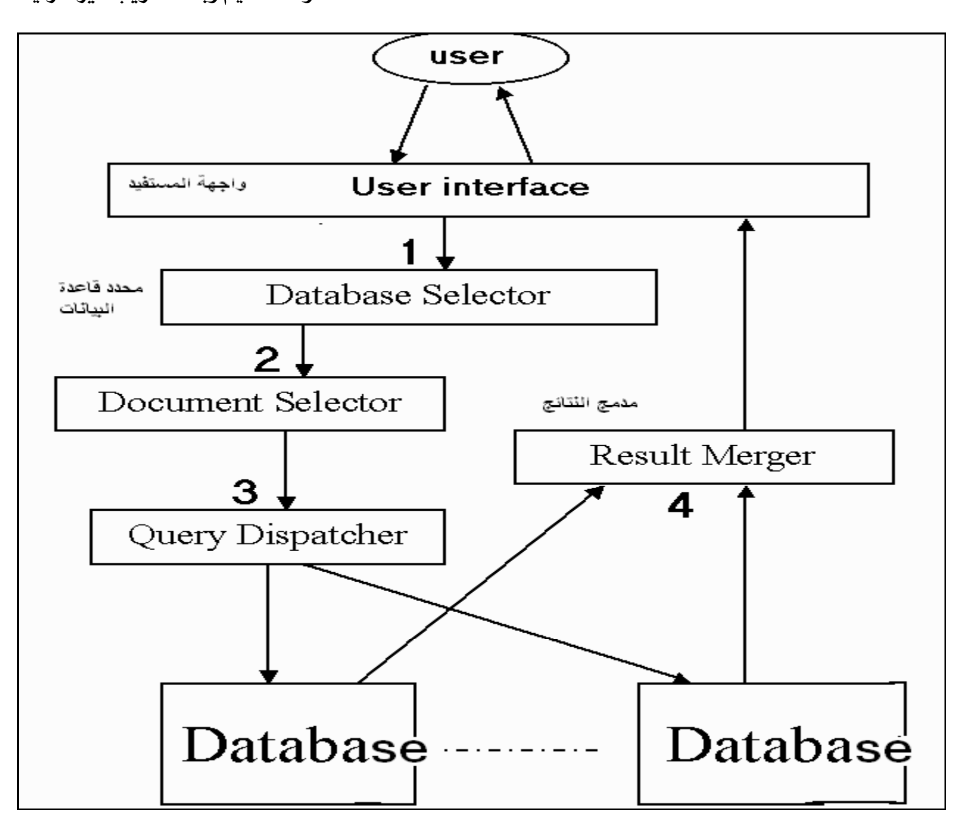
٢- تستطيع محركات بحث الويب غير المرئية أن تفتح أمام مستخدمي الويب منفذًا جامعًا لمختلف تخصصات قواعد البيانات، لأن طبيعة التعامل مع قاعدة البيانات تكون من خلال نموذج وبنية البحث، وليس من خلال المحتوى أو الموضوع.

(١) Meng, Weiyi and Yu, Clement. Building efcient and eective metasearch engines, State University of New York, , cited 28/8/2007, cited at http://portal.acm.org/citation.cfm?id=505284

٣- يمكن لمحركات بحث الويب غير المرئية أن تجمع شتات مصادر الموضوع الواحد بطرح سؤال واحد وتلقي النتائج المتنوعة على شاشة واحدة، وذلك عندما تنتهج عمل محركات البحث المتعددة في بحث أكثر من قاعدة بيانات في وقت واحد.

١/٢/٣/٣ مكونات عمل البحث المتعدد مع قواعد البيانات:

تعتمد محركات البحث المتعددة على أجزاء متكاملة يؤدي كل منها مهام محددة في بحث النتائج واسترجاعها من قواعد البيانات. تبدأ هذه المهام منذ تلقي محرك البحث استفسار المستفيد ثم تحديد الاتجاهات الموضوعية لاستفسار البحث واختيار قواعد البيانات ذات التوافق الموضوعي مع كلمات البحث، ثم طرح الاستفسار على القواعد متلقيًا النتائج المستدعاة وإعادة ترتيبها مرة أخرى لكي تلائم العرض مجتمعة على نافذة واحدة أمام مستخدم الويب. ويمثل الشكل (٣-٦) مكونات عمل البحث المتعدد مع قواعد البيانات، وقد وُضعت هذه المكونات في سياق الإجراءات التفاعلية بين محرك البحث وقواعد البيانات في طرح الاستفسار وتلقي الإجابة.



الشكل (٣-٦) مكونات عمل محركات البحث المتعددة للويب غير المرئية

يوضح الشكل (٣-٦) مكونات عمل محركات البحث المتعددة ؛ حيث يدل كل رقم في الشكل السابق على أحد المكونات مرتبة في سياق الإجراءات والمهام التي يؤديها محرك البحث المتعدد. وهي كالتالي(١):

(١) **محدد قاعدة البيانات database selector** : يعمل محدد قواعد البيانات على تحقيق التوافق بين موضوع البحث الممثل في مصطلحات استفسار البحث وبين قواعد البيانات التي تخدم الموضوع نفسه دون غيرها من قواعد البيانات الأخرى، وذلك باستخدام عملية المضاهاة. ويعتمد محدد القواعد في الانتقاء والاختيار على الكلمات الدالة المعبرة عن كل قاعدة بيانات على حدة، إضافة

(١) Meng, Weiyi and Yu, Clement Ibid

إلى البيانات الإحصائية المعبرة عن قيمة قاعدة البيانات في تلبية حاجات موضوعية محددة. ويمكن القول إنه كلما زاد عدد قواعد البيانات وزادت موضوعاتها وتخصصاتها علت أهمية عمل محدد قواعد البيانات، فالتحكم في عدد صغير من قواعد البيانات أمر يسير خلاف ما يمكن أن يعانيه مستخدم الويب من استدعاء وثائق لا علاقة لها باستفسار البحث، وهذا ما يمكن أن يترتب على الأعداد الكبيرة من قواعد البيانات دون استخدام محدد القواعد. وعليه، فإن عملية المضاهاة أو المقابلة التي يقوم بها محدد القواعد تجنب كلاً من محرك البحث المتعدد ومستخدم الويب كثيرا من المشكلات التي تكمن فيما يلي:

أ- عند غياب محدد القواعد لا يستطيع محرك البحث تحديد قواعد البيانات الملائمة للبحث ومن ثم كان البحث في كل قواعد البيانات دون النظر إلى التخصص، مما يستدعي معه وثائق لا علاقة لها بالموضوع.

ب- تثير النتائج المستدعاة كمًا هائلاً من البلبلة لا يمكن لمستخدم الويب الحصول منها على يريده من المعلومات في وقت قصير.

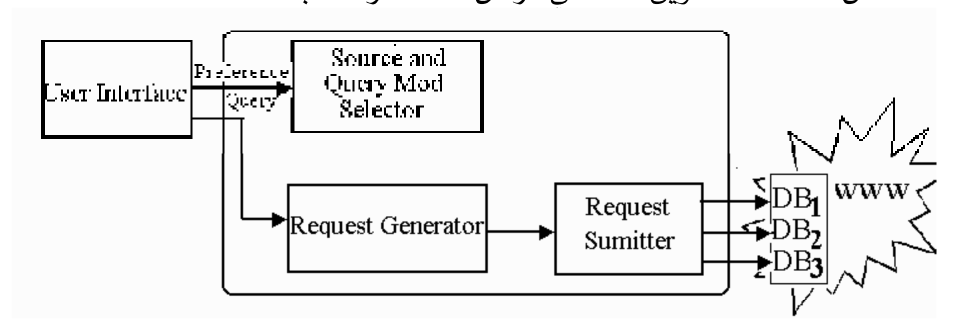
ج- البحث في كل قواعد البيانات مع زيادة أعدادها يفقد محرك البحث الدقة والكفاءة في استرجاع النتائج الملائمة.

(٢) محدد الوثائق (التسجيلات) document selector : يقوم محدد الوثائق بالتعرف إلى مجموعة الوثائق ذات الأهمية الكبرى في الموضوع ثم استرجاعها دون غيرها من كل وثائق قاعدة البيانات. مما يوفر معه الوصول إلى أعلى درجة تحقيق وأدق نسبة استدعاء أمام مستخدم الويب. وثمة عوامل تتحكم في عمل محدد الوثائق مثل مدى التشابه في الموضوع بين الوحدات المسترجعة، ومدى التخصص الموضوعي لقاعدة البيانات. هذا بجانب مراعاة العوامل التالية:

- **احتياجات المستخدم** : حيث يقدم محرك البحث خدماته لكل مستخدمي الويب على المستوى العالمي، إلى جانب تعامله مع قواعد البيانات المختلفة بلا تحديد.

- **تفاوت الأهمية للوثائق:** حيث تتفاوت مجموعات الوثائق المستدعاة فيما بينها في مدى الصلة بالموضوع.
- **خلفية التعامل مع قواعد البيانات:** حيث يعكس تاريخ عمليات البحث السابقة أهمية الوثائق ووزنها الموضوعي في تلبية الاحتياجات الموضوعية المختلفة.
- **تحقيق الشمولية في الاستدعاء :** وهو المبدأ الذي يقتضي استرجاع مختلف الوثائق ذات الصلة القوية أو الضعيفة بالموضوع.

(٣) مرسل استفسار البحث: Query dispatcher : يقوم مرسل الاستفسار ببناء اتصال بين محرك البحث المتعدد وقواعد محركات البحث الأخرى كي يطرح عليها استفسار البحث، ويتم هذا الاتصال باستخدام قناة بروتوكول النص الفائق HTTP للاتصال. والجدير بالذكر أن مرسل الاستفسار يتحرى ترجمة وصياغة استفسار البحث بما يلائم لغة البحث داخل كل قاعدة بيانات، فمن الخطأ إرسال استفسار البحث بلغة واحدة عامة لكل قواعد البيانات المختلفة أو بتلك التي يفهمها محرك البحث المتعدد ذاته. وثمة أمور يجب الاهتمام بها مثل إعادة وزن الأهمية لمختلف كلمات البحث حتى يمكن تحديد قواعد البيانات الأكثر ملاءمة لإخراج النتائج، أيضًا الاتجاهات الموضوعية لمستخدم الويب واحتياجاته من الوثائق التي يجب استدعاؤها أو التي يمكن إهمالها. ويوضح الشكل (٣-٧) التكوين الداخلى لمرسل استفسارات البحث.



الشكل (٣-٧) التكوين الداخلي لمرسل استفسارات البحث

(٤) دمج وصياغة النتائج (المدمج): Result merger : تأتي النتائج المستدعاة من مختلف قواعد البيانات وقد حملت كل مجموعة منها خصائص قاعدة البيانات (المصدر)، وتتمثل هذه الخصائص في الصياغة والترتيب. ويأتي دور برنامج دمج النتائج للجمع بين الملامح المختلفة وإعادة ترتيبها وصياغتها في ترتيب نسق واحد يلائم العرض على واجهة النتائج الرئيسية لمحرك البحث المتعدد أمام مستخدم الويب. ويستخدم مدمج النتائج آليات مختلفة لوزن الأهمية وترتيب النتائج حتى تكون نهاية قائمة موحدة لكل النتائج مرتبة من الأكثر إلى الأقل أهمية. ومن بين الاستراتيجيات التي يعتمد عليها محرك البحث المتعدد في وزن وترتيب النتائج المختلفة ما يلي[1]:

١- إعطاء الوزن والأهمية الأكبر لأوائل النتائج : حيث يرى مدمج البيانات أن النتائج المرتبة ذات الأعداد (١- ن) هي نتائج مهمة ومن ثم تأخذ أعلى قيم ترتيب في قائمة النتائج العامة.

٢- حساب مدى الصلة بين النتائج وكلمات استفسار البحث : حيث تحدد أهمية الكلمات الواردة في النتائج ومطابقتها مع أهمية الكلمات الواردة في استفسار البحث.

٣- إعطاء الوزن والأهمية الأكبر لأوائل التسجيلات المسترجعة : وهو ما يشبه استراتيجية الدمج الأولى غير أنها لا تتعامل مع الوثائق وإنما تتعامل مع التسجيلات المخصصة للوثائق داخل قواعد البيانات.

٢/٢/٣/٣ آلية عمل البحث المتعدد مع قواعد البيانات

تتكامل مكونات محرك البحث المتعدد السابق ذكرها لتؤدي كل منها وظائف محددة في منظومة العمل مع قواعد البيانات. وتعتمد آلية العمل على ثلاثة برامج فرعية هي:

(١) Lu, Yiyao and Meng, Weiyi. Evaluation of result merging strategies for metasearch engines, University of Illinois at Chicago, 2005, cited 20/10/2007, cited at http://www.cs.binghamton.edu/~meng/pub.d/Lu_p211.pdf

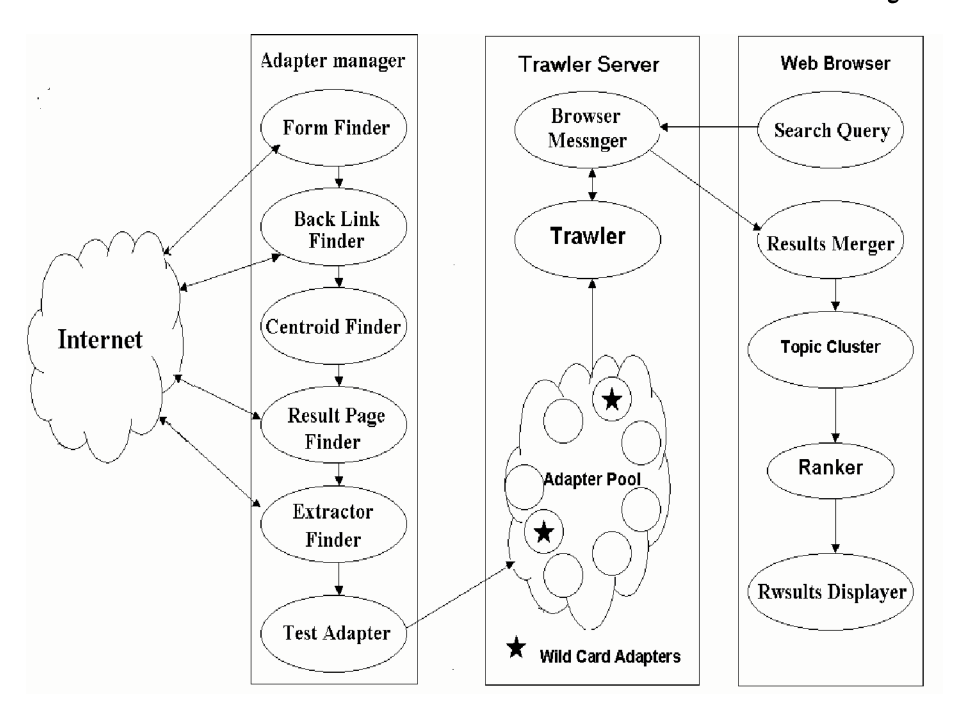
أولاً: مدير التهيئة Adapter Manager:

ثانيًا: خادم المحرك Trawler Server:

وثالثًا: المتصفح Browser:

تعمل هذه البرامج على توفير مهام الوصول إلى قاعدة أو قواعد البيانات، ثم تحديد وحدات البيانات ذات الصلة بالموضوع، وأخيرا استرجاع هذه الوحدات إلى نافذة مستخدم الويب. ويتضح عمل كل منها من خلال التعرف إلى ميكانيكية عمل محرك البحث Turbo10. وتمكن هذه البرامج محرك البحث من تحقيق كفاءة أعلى وسرعة أكبر؛ حيث يتم الاتصال مع قواعد البيانات وطرح الاستفسارات عليها وتلقي الإجابة في شكل مجموعات من نتائج البحث تحتاج إلى إعادة ترتيب وتهيئة مرة أخرى لتصاغ في قائمة واحدة مرتبة، غير أن عمليات التهيئة والمعالجة لهذه النتائج لا تتم داخل خادم محرك البحث، وإنما يرسلها الخادم في شكل رسائل غير متزامنة إلى متصفح مستخدم الويب نفسه، وذلك لتتجمع في الذاكرة المخصصة للمتصفح في جهاز المستخدم، وهذه الذاكرة تتلقى النتائج ثم معالجتها دون المرور بالمعالجة داخل خادم محرك البحث. هذا إلى جانب إمكانات التقسيم الموضوعي التي يتميز بها محرك البحث عند استخدام هذه البرامج. ويوضح الشكل(٣-٨) آلية عمل هذه البرامج داخل محرك Turbo10 [١].

(١) Hamilton, Nigel. the mechanics of a deep net metasearch engine, Turbo10 Metasearch Engine, 2004, cited 15/10/2007, cited at
http://www2003.org/cdrom/papers/poster/p170/poster/poster.html



الشكل (٣-٨) آلية العمل لبرامج محرك Turbo10

يوضح الشكل (٣-٨) البنية العامة وآلية العمل المكونة من برامج ثلاث ممثلة في ثلاثة أعمدة يحوي كل منها عدة عناصر للعمل هي كالتالي:

أولاً: مدير التهيئة : Adapter manager يقتضي واقع عمل محرك البحث مع قواعد بيانات الويب التعامل مع أشكال ونماذج واجهات بحث مختلفة التصميم والبنية، ويمثل مدير التهيئة هنا مسئول التخاطب بين محرك البحث وقواعد البيانات، حيث يحدد شكل وبنية القاعدة ثم إقامة العلاقة إرسالاً واستقبالاً مع قاعدة البيانات. ويبدأ عمل مدير التهيئة بشكل يدوي في تحديد رابط قاعدة البيانات URL من جانب العنصر البشري، ثم يستمر التعامل آليًا أو ذاتيًا في ترجمة وتهيئة استفسار البحث للإرسال ثم ترجمة وتهيئة نتائج البحث للاستقبال والعرض أمام مستخدم الويب.

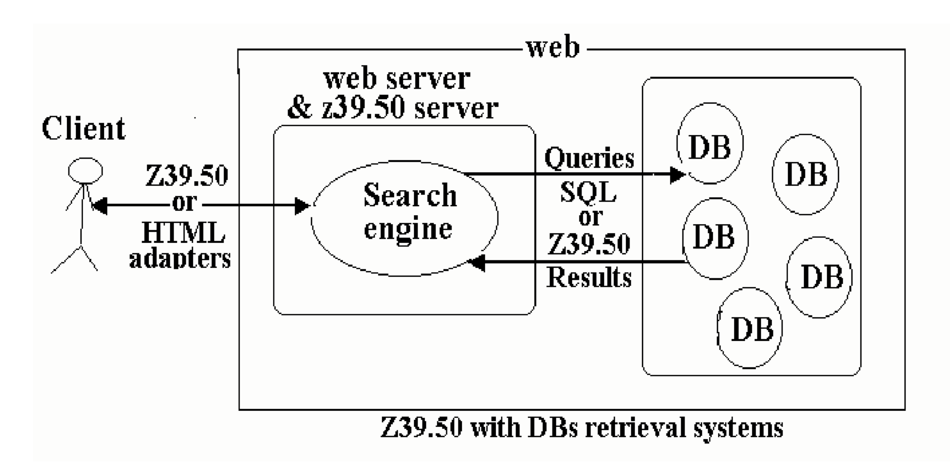
ويتكون مدير التهيئة من عناصر فرعية لتسيير إجراءات التحاور مع قواعد البيانات تبدأ باستخدام عنصر form finder لاستخدام روابط الويب في البحث عن نماذج واجهات البحث وتحليلها لاكتشاف خصائص بناء النموذج والتصميم في كل واجهة أو نموذج بحث. ينتهي هذا الإجراء عند عنصر Back Link Finder لكي تطرح استفسارات الاختبار واسترجاع النتائج لتحليلها ومن ثم تكشيف وتحديد الموضوع أو المجال الذي تهتم به قاعدة البيانات. أما عنصر Centroid finder فإنه يعمل على تحليل واقتباس الكلمات ذات الدلالة الموضوعية ووزن أهمية هذه الكلمات لتحديد الكلمات الأكثر توضيحًا لموضوع قاعدة البيانات حيث تستخدم هذه الكلمات فيما بعد أيضًا لتشكيل استفسارات اختبار المحتوى لقواعد البيانات test queries، عند استرجاع نتائج البحث من قواعد البنات فإن Result Page Finder يعمل على تلقي هذه الإجابات وتحليلها بواسطة Extractor Finder الذي يعمل على تحديد النتائج ذات الصلة باستفسار البحث ويتميز عنصر الاقتباس هنا بإمكانية تحليل صفحات النتائج الواردة من قواعد بيانات مختلفة في وقت واحد. وينتهي عمل مدير التهيئة عند إرسال النتائج المحددة لخادم المحرك trawler server.

ثانيًا: خادم المحرك Trawler server : ويعد هذا البرنامج واسطة المعالجة الحقيقية لكل من استفسارات البحث الواردة من مستخدمي الويب، ونتائج البحث الواردة من قواعد البيانات. ويعتمد خادم المحرك في عمله على the pool of adapters. ويحوي هذا العنصر بداخله مجموعة من Wild Card Adapters، تقوم هذه المعاملات بصياغة التوافق وإجراء المضاهاة بين كلمات البحث وقواعد البيانات الملائمة للبحث حتى توجه استفسارات البحث إلى القاعدة أو القواعد الملائمة. وترجع قدرة العمل في المضاهاة على تقسيم موضوعي وتكشيف لمجموعات قواعد البيانات التي يتعامل معها محرك بحث قواعد بيانات الويب غير المرئية.

ثالثًا: متصفح الويب :Web browser : ويمثل هذا البرنامج الواجهة العامة لمحرك البحث المسؤولة عن تلقي استفسارات البحث من جانب مستخدمي الويب وإرسالها إلى المعالجة والمضاهاة مع موضوعات قواعد البيانات. إلى جانب قيامها بتلقي الإجابات المعالجة من خادم المحرك وعرضها أمام مستخدم الويب.

٣/٣/٣ معايير بحث نظم قواعد البيانات:

تمثل معايير البحث هنا لغة الحوار والتخاطب بين محركات البحث وقواعد البيانات ؛ ذلك لما يتسم به كل طرف من استخدام لغات وآليات العمل والبحث الخاصة به، فضلاً عن اختلاف قواعد البيانات فيما بينها من حيث نظم الإدارة ولغات البحث؛ مما أوجب استخدام لغة بحث واحدة تستطيع العمل مع نماذج مختلفة من واجهات البحث. ولا يقصد الباحث هنا الحديث عن معيار بعينه وإنما الحديث عن معايير البحث كلغة وسيطة بين طرفي محرك البحث في طرح الاستفسار وقواعد البيانات وإعطاء النتائج. وإذا ما خلا التعامل بين محركات البحث وقواعد البيانات من معيار موحد للبحث، فلا يمكن لقواعد البيانات التعامل مع الاستفسارات الواردة من محرك البحث، ومن ثم يصعب تلقي نتائج بحث لاستفسارات غير مفهومة. ويأخذ معيار البحث مكانه كرابط بين أضلاع مثلث استرجاع محتوى قواعد البيانات ؛ حيث يبدأ عمله عند مستخدم الويب الذي يتعامل مع واجهة بحث مهمة مصوغة باستخدام معيار البحث، ثم الدور الوسيط الذي يقوم به محرك بحث قواعد البيانات في تحديد قاعدة أو قواعد البيانات الملائمة لاستفسار البحث، وينتهي عمل معيار البحث عن قواعد البيانات التي تستخدم هذا المعيار للتعامل مع لغات بحثها المختلفة في فهم السؤال واستدعاء النتائج. وأهم معايير البحث المستخدمة لذلك معيارا SQL, Z39.50. ويوضح الشكل (٣-٩) مواضع عمل معايير البحث في نظم استرجاع قواعد البيانات.



الشكل (٣-٩) مواضع عمل معايير البحث في نظم استرجاع قواعد البيانات

يمثل الشكل (٣-٩) خريطة مبسطة للكشف عن مواقع عمل معايير البحث داخل محرك بحث الويب غير المرئية. وسوف يتم التركيز على خصائص معيار البحث Z39.50 وخصائص تفعيله داخل محرك البحث المقترح في الفصل السادس من هذه الدراسة.

٣/ ٤ خاتمة:

تناول هذا الفصل جوانب المعالجة الفنية والتنظيم لقواعد بيانات الويب ؛ حيث عكست الدراسة أن قواعد بيانات الويب يتم تناولها من خلال وجهتي نظر ؛ تتعلق إحداهما بالتعامل مع قواعد البيانات كونها نظمًا لاسترجاع المعلومات يمكن من خلالها البحث والاسترجاع لملفات المعلومات الرقمية على الويب، أما وجهة النظر الثانية فتتعلق بالتعامل مع قاعدة البيانات كونها مصدرًا للمعلومات يمكن العمل على تنظيمه ووصفه فنيًا في مجموعة من حقول الوصف الفني حتى يتم التعبير عن ماهية ومحتوى هذه القواعد. وقد تناول هذا الفصل محركات بحث الويب غير المرئية كونها تنطابق تمامًا في عملها مع الآلية التي تسير عليها محركات البحث

المتعددة، ومن ثم فإن آلية العمل هنا يمكن تطويعها للتعامل مع تنظيم واسترجاع ملفات قواعد الويب غير المرئية. كما تبنى الفصل أيضًا فكرة النظر إلى أدوات البحث المقترحة لبحث الويب غير المرئية على أنها ليست محرك بحث فقط، وإنما بوابة كبيرة لاسترجاع ملفات قواعد البيانات، تستطيع تقديم الكثير من الخدمات الأخرى المتعلقة بقواعد البيانات والتعامل معها.

الفصل الرابع

واقع قواعد بيانات الويب غير المرئية

الدراسة التجريبية

٠/٤ تمهيد :

خلصت الفصول السابقة من الدراسة إلى التعرف إلى أسس عمل قواعد بيانات الويب وآليات التنظيم والمعالجة المنتظرة لقواعد البيانات داخل محركات بحث الويب غير المرئية، وقد ركز الفصل الثاني خاصة على واقع قواعد البيانات وامتزاجها مع بيئة الويب من جانب والآليات المتبعة لنظم إدارة وتشغيل قواعد البيانات من جانب آخر، كما ركز الفصل الثالث على أدوات البحث. ويركز هذا الفصل على الواقع التطبيقي لعينة قصدية مختارة من قواعد البيانات على الويب، تجمع بين قواعد البيانات النصية وقواعد بيانات الوسائط المتعددة ؛ يتم رصدها من حيث نماذج البحث وعناصرها المتواجدة في واجهات البحث، أيضا خصائص الاستدعاء والمعالجة للوحدات أو التسجيلات التي تحويها قواعد بيانات الويب، والهدف من ذلك هو التعرف إلى الخصائص الواقعية للعمل داخل قواعد بيانات الويب حتى تصاغ هذه الخصائص في شكل ملامح لبناء وهيكلة قواعد البيانات، وهي التي سينبني عليها عمل محركات بحث قواعد بيانات الويب (نظم استرجاع قواعد البيانات).

١/٤ خصائص عمل قواعد البيانات على الويب :

يشتمل بناء وعمل قواعد بيانات الويب على مجموعة كبيرة من الخصائص تتنوع بين الخصائص التقنية لنظم إدارة قواعد البيانات، وبناء الجداول وتنظيم التسجيلات، وبين الخصائص المعلوماتية في نماذج البحث الحر أو التصفح والبحث بالحقول إلى جانب جوانب الاسترجاع والتنظيم للنتائج المستدعاة من قواعد البيانات. وقد اختار الباحث مجموعة الخصائص المؤثرة في عمل وكفاءة محركات بحث قواعد البيانات فيما بعد دون الخصائص الأخرى التي لا تتقاطع مساراتها مع آليات عمل محركات البحث. وجدير بالذكر هنا أن هذه

الخصائص تنسحب على مختلف قواعد البيانات وإن اختلفت في شكل المحتوى بين النصي والوسائط المتعددة، إلى جانب تساوي قواعد البيانات النصية والببليوجرافية وقواعد المستخلصات، وهذا يرجع إلى أن قواعد البيانات جميعها تشترك في بناء التسجيلات اعتمادًا على الجداول والحقول، وتختلف فقط في إضافة حقل أو عنصر بحث يمكنه التعامل مع المواد غير النصية من صوت وصورة تخضع دائمًا لآليات التحليل المبني على المحتوى CBR. ويبني تحليل هذه الخصائص على الملاحظة والمعايشة لحالة قواعد البيانات من واقع طرح الاستفسارات البحثية واسترجاع النتائج.

ولقد حرص الباحث على تمثيل واقع قواعد بيانات العينة من خلال تقسيم الخصائص موضع الدراسة إلى قسمين أساسيين ؛ أحدهما : يختص بدراسة الملامح أو الخصائص العامة التي تنسحب على نوعي قواعد البيانات النصية المتخصصة، وقواعد بيانات الوسائط المتعددة العامة. أما الآخر : فيمثل مجموعة الخصائص النوعية التي يتميز في تطبيقها نوع من قواعد البيانات النصية المتخصصة أو الوسائط العامة على الآخر. ومن ثم يتضح التقسيم في الشكل التالي:

أولاً: الخصائص العامة :

٤/١/١ واجهة البحث.

٤/١/٢ أساليب البحث.

٤/١/٣ مستويات الوصول.

٤/١/٤ خدمات قواعد البيانات.

ثانيًا: الخصائص النوعية :

٤/١/٥ واجهات وأنماط عرض النتائج.

٦/١/٤ محتوى قواعد البيانات.

٧/١/٤ عرض التسجيلات النهائية.

ويمكن تناول هذه الخصائص كالتالي:

الخصائص العامة لقواعد البيانات:

١/١/٤ واجهة البحث:

بلغت أهمية واجهات بحث قواعد البيانات القدر الذي جعل أخصائيي المعلومات والهيئات القائمة على أنشطة المكتبات العمل على إيجاد خطوط إرشادية للاختيار والتعامل مع واجهات بحث قواعد بيانات الويب. وحرصت هيئة مصادر المعلومات الإلكترونية CAUL Electronic Information Resources Committee (CEIRC) على إصدار بعض الملامح العامة الخاصة بواجهات بحث قواعد البيانات، التي من خلالها تستطيع المكتبات المفاضلة بين قواعد البيانات بعضها بين بعض، ومن بين هذه الملامح الإرشادية ما يلي[1]:

١- واجهة قاعدة البيانات Database Entry Page.

٢- البحث Searching.

٣- تقييد البحث Limiting.

٤- الرسومات Graphics.

٥- عرض التسجيلات Viewing Records.

٦- استرجاع التسجيلات Record Retrieval.

(١) Woodcroft, Michelle and Fletcher, Janet. Establishing guidelines for the effective evaluation of web-based periodical bibliographic and full-text databases search interfaces, Southern Cross University, 2000, cited 1/11/2007, cited at www.vala.org.au/vala2000/2000pdf/Woo_Fle.pdf

٧- تقديم التسجيلات Record Output.

٨- البث الانتقائي للمعلومات Selective Dissemination of Information (SDI).

٩- التصميم العام General Design.

١٠- المصطلحات Terminology.

١١- المساعدة Help.

١٢- الملامح المتقدمة Advanced Features.

غير أن الباحث قد عمل على توزيع الملامح السابقة على الخصائص التالية في دراسة واقع قواعد البيانات، بحيث توزع على عناصر مختلفة تدرس بحسب القرب أو البعد في تشابه العمل. وكانت عناصر واجهة البحث كالتالي:

١/١/١/٤ واجهة مباشرة أو مركبة:

تستخدم نظم استرجاع البيانات عامة شكلين من أشكال البحث المقدمة لمستخدمي النظام تتمثل في الواجهة المباشرة للبحث وهي التي تحمل عنوان النظام ومختلف عناصر البحث المقدمة لبحث التسجيلات الداخلية، أما الشكل الثاني فهي واجهة البحث المركبة أو المتقدمة المتمثلة في عناصر أكثر تحديدًا وتنقيحًا للبحث، ويتم الوصول إليها من خلال رابط في الواجهة الأساسية تحت عنوان بحث متقدم.

الجدول (٤-١ / أ) طبيعة واجهات البحث في قواعد البيانات النصية

واجهة البحث	قواعد بيانات نصية متخصصة				
	السيرة	التربية	الحيوان	الطب	التسوق
	biography	**eric**	**Animal search**	**medlineplus**	**amazon**
مباشرة	√	- - -	√	√	√
مركبة (متقدمة)	- - -	√	- - -	- - -	- - -

الجدول (٤-١ / ب) طبيعة واجهات البحث في قواعد بيانات الوسائط المتعددة

واجهة البحث	قواعد بيانات وسائط متعددة					
	الصوت		صور ثابتة		صور متحركة	
	allmusic	rhapsody	corbis	freefoto	movies.go	world-of-movies
مباشرة	- - -	√	- -	√	√	√
مركبة(متقدمة)	√	- - -	√	- -	- - -	- - -

توضح الجداول السابقة (٤-١ / أ) و (٤-١ / ب) حالة واجهات البحث داخل قواعد بيانات العينة النصية والوسائط المتعددة، وقد عكست هذه الجداول ما يلي:

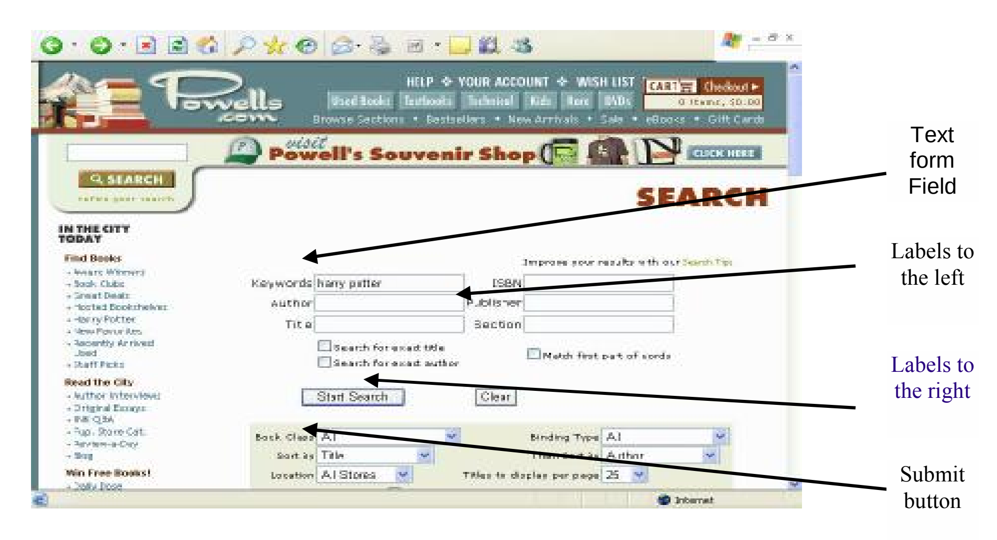
أ- **القواعد النصية:** اعتمدت قاعدة Eric على استخدام واجهات إضافية تتميز بعناصر بحث متعددة ومتشعبة يمكنها تنقيح وتدقيق إستراتيجية بحث مستخدم الويب تحت عنوان Advanced Search، وهذا إنما يرجع للطبيعة شبه الأكاديمية التي تتبناها قاعدة البحث دون القواعد الأخرى التي تتسم بعمومية الاستخدام. غير أن قاعدة البيانات medlineplus التي تنتهج أيضا المنحى الأكاديمي قد اعتمدت على واجهة مباشرة واحدة بما لا يتناسب مع تشعب واتساع المجال الموضوعي الذي تغطيه.

ب- **قواعد الوسائط:** اعتمدت قاعدتا corbis, allmusic على إتاحة البحث المتقدم في بحث مجموعاتها؛ وذلك يرجع إلى العامل المشترك بين القاعدتين في كبر حجم مجموعاتها مما يساعد على زيادة الاستدعاء بشكل كبير في حالة البحث العام، ولذلك فقد عملتا على توفير عناصر مختلفة متعددة للبحث تمكن عامل التحقيق من الارتفاع وانخفاض عامل الاستدعاء. ويأتي استخدام قواعد بيانات الوسائط المتعددة لواجهة البحث المباشرة نتيجة لسهولة وسرعة البحث من خلالها، خلاف واجهات البحث

المتقدمة فتحتاج الى معرفة ودقة أكبر. ويرى الباحث أن اعتماد قواعد البيانات على واجهات إضافية متقدمة للبحث لابد أن يؤخذ في الاعتبار عند بناء نظام استرجاع قواعد البيانات، وخاصة في عمل برنامج الزاحف الذي عليه أن يراجع الواجهات التي تحتوي على نماذج وعناصر البحث عند إضافة أو استبعاد قواعد البيانات. إلى جانب ذلك تزويده بآليات تحليل روابط الواجهة الأساسية لقواعد البيانات والبحث عن رابط البحث المتقدم advanced search أو ما يرادفه من مصطلحات أخرى.

٢/١/١/٤ نموذج البحث:

يحوي نموذج البحث مجموعة العناصر التي تشكل استراتيجية البحث لمجموعات قاعدة البيانات، وتعبر مختلف عناصر النموذج الواحد عن الخصائص والوصف العام لمجموعات البيانات، وغالبًا ما تكون هذه العناصر هي مسميات حقول التسجيلات الداخلية لمحتوى قاعدة البيانات. ويشتمل نموذج البحث على أقسام فرعية يحتوي كل منها على عناصر البحث المتشابهة في الخصائص، كأن يوجد قسم لبحث بيان المسؤولية الفكرية يتكون من عناصر هي الاسم الأول واللقب والبحث بالمترجم. أيضًا قسم بيانات النشر الذي يحتوي على عناصر مكان النشر والناشر وتاريخ النشر. وكما سبقت الإشارة، فإن واجهات بحث قواعد البيانات تستخدم إما النموذج العام للبحث والمكون من فراغ البحث لإدخال الكلمات المفتاحية، أو أن تستخدم النماذج المتعددة المعتمدة على عناصر مختلفة لبحث المحتوى.



الشكل (٤– ١) مثال لنموذج بحث مواقع الويب

الجدول (٤–٢ / أ) نموذج بحث واجهة قواعد البيانات النصية

نموذج البحث		قواعد نصية متخصصة				
		السيرة	التربية	الحيوان	الطب	التسوق
		biography	eric	Animal search	medlineplus	amazon
نموذج عام		√	√	√	√	√
نموذج متعدد	فراغ النص	- - -	√	- - -	- - -	- - -
	زر الاختيار	- - -	√	- - -	- - -	- - -
	زر الخصائص	- - -	√	- - -	- - -	- - -
	قائمة الاختيارات	- - -	√	- - -	- - -	- - -
	عنصر الترتيب والعرض	- - -	- - -	- - -	- - -	- - -

الجدول (٤-٢ / ب) نموذج بحث واجهة قواعد بيانات الوسائط المتعددة

نموذج البحث		قواعد وسائط متعددة					
		الصوت		صور ثابتة		صور متحركة	
		allmusic	rhapsody	corbis	freefoto	movies.go	world-of-movies
نموذج عام		- - -	√	- - -	√	√	√
نموذج متعدد	فراغ النص	√	- - -	√	- - -	- - -	- - -
	زر الاختيار		- - -	√	- - -	- - -	- - -
	زر الخصائص	√	- - -	√	- - -	- - -	- - -
	قائمة الاختيارات	√	- - -	√	- - -	- - -	- - -
	عنصر الترتيب والعرض	√	- - -	√	- - -	- - -	- - -

من الجداول السابقة يتضح :

أن أغلب قواعد البيانات استخدمت النماذج العامة للبحث معتمدة على عنصر أو اثنين على الأكثر في عمليات بحث واستدعاء الوحدات الداخلية لمجموعات قواعد البيانات. وقد تكون نموذج البحث العام في هذه القواعد من عنصر كتابة الكلمات المفتاحية إلى جانب استخدام قائمة الاختيارات لتحديد فئة أو شكل مصدر المعلومات المراد البحث عنه. وترجع الفكرة الأساسية في ذلك إلى المفهوم السائد لدى مصممي نظم الاسترجاع على الويب في التعامل مع المستوى الأدنى من مهارات البحث لدى مستخدمي الويب.

٣/١/١/٤ عناصر البحث:

تتكون نماذج البحث من مجموعة من عناصر البحث يختص كل منها بإدخال قيمة محددة يتم البحث بها، وتتكامل هذه القيم المدخلة بواسطة عناصر البحث لتحديد ملامح وخصائص الوحدات أو النتائج المراد البحث عنها. ويتكون كل عنصر من عناصر البحث من شقين؛ الأول: عنوان العنصر attribute label، والثاني: العنصر ذاته attribute element. وحري بالذكر أن هناك بعض العناصر التي تتكون من عناصر أخرى فرعية تتيح للمستخدم الصياغة الأمثل لاستراتيجية البحث مثل الاختيار بين البحث بالاسم الأول أو الأخير أو اللقب في حقل المؤلف. ولاختلاف طبيعة الحقول وعناصر الوصف المعبرة عن الوحدات الداخلية لمحتوى قواعد البيانات إلى جانب اختلاف طبيعة وأشكال مصادر المعلومات ذاتها، فإن عناصر البحث تختلف فيما بينها في الشكل وطبيعة الأداء في عملية البحث. ومن بين أنواع عناصر البحث ما يلي[١]:

١- فراغ الكلمات البحثية text box.

٢- زر الاختيار radio button.

٣- زر الخصائص checkbox.

٤- قائمة الاختيار selection list.

ويضيف الباحث عنصرًا آخر، هو:

٥- تحديد ملامح النتائج: وهو الخاص بعدد النتائج المستدعاة في الصفحة الواحدة أو إظهار وإخفاء بعض جوانب الوحدات المستدعاة.

يوضح الجدولان (٤-٢ / أ) و (٤-٢ / ب) السابقان أنه من بين قواعد بيانات العينة استخدمت ثلاث قواعد فقط نموذج البحث المتعدد والمكون من مجموعة

(١) He, Hai and Yu, Clement. Automatic Extraction of Web Search Interfaces for Interface Schema Integration, 2002, cited 11/11/2007, cited at www.cs.binghamton.edu/~meng/pub.d/WWWposterhe.pdf

من عناصر البحث المختلفة، وهذه القواعد هي eric, Allmusic and corbis ، وهو ما جاء نتاجًا لعمومية المجالات الموضوعية لقواعد البيانات أو اتجاهها لخدمة مختلف مستويات التخصص والمعرفة بالمجال من مستخدمي الويب. وقد احتوت قاعدة corbis على مختلف أشكال عناصر البحث، وهو ما جاء من تعدد عناصر وصف الصور التي يمكن البحث والاسترجاع من خلالها. وينطبق ذلك على قاعدة allmusic غير أنها لم تحتوِ على عنصر الاختيار radio button. وجدير بالذكر هنا احتواء هاتين القاعدتين على عنصر للبحث يختص بطرق التنظيم أو العرض لمجموعات الوحدات المسترجعة، وقد ورد هذا العنصر بعنوان Point of View في قاعدة البيانات corbis.

ويرتبط نموذج البحث وعناصره بالتعريف بقاعدة البيانات أمام برامج الزاحف ومحركات بحث قواعد البيانات ؛ حيث سيرتبط قرار الزاحف بضم أو استبعاد واجهات بحث قواعد البيانات بما يجده من نماذج وعناصر البحث داخل هذه الواجهات. ومن ثم تحتاج واجهات ونماذج البحث إلى التوحيد والتقييس في شقي العنونة والصياغة البرمجية للغات البحث. إلى جانب ذلك فإن تعدد عناصر البحث واختلافها يمكن أن يؤثر إيجابيًا في إرسال الاستفسارات من محركات البحث لقواعد البيانات ؛ فكلما كثرت عناصر البحث كلمات استطاعت قواعد البيانات فهم الاستفسارات واسترجاع النتائج الموافقة لها.

٢/١/٤ أساليب البحث:

تستخدم قواعد بيانات الويب أسلوبين لإجراء عمليات البحث، هما : أسلوب البحث بالتصفح، وأسلوب البحث الحر. وتمثل أساليب البحث النوافذ الأساسية لاسترجاع الوحدات الداخلية لمحتوى قواعد البيانات، ويختلف قرار قاعدة البيانات بتبني أحد الأسلوبين اعتمادًا على طبيعة المحتوى والمجال الموضوعي الذي تقدمه قواعد البيانات. ويوضح الجدولان التاليان أنواع أساليب البحث داخل قواعد بيانات العينة:

الجدول (٤-٣ / أ) أساليب البحث في قواعد البيانات النصية

أساليب البحث	قواعد نصية متخصصة				
	السيرة	التربية	الحيوان	الطب	التسوق
	biography	eric	Animal search	medlineplus	amazon
البحث بالتصفح	√	√	√	√	√
البحث الحر	√	√	√	√	√

الجدول (٤-٣ / ب) أساليب البحث في قواعد بيانات الوسائط المتعددة

أساليب البحث	قواعد وسائط متعددة					
	الصوت		صور ثابتة		صور متحركة	
	allmusic	rhapsody	corbis	freefoto	movies.go	world-of-movies
البحث بالتصفح	- - -	- - -	- - -	√	- - -	- - -
البحث الحر	√	√	√	√	√	√

١/٢/١/٤ البحث بالتصفح:

تشير الجداول (٤-٣ / أ) و (٤-٣ / ب) إلى مجموعة من الملامح المتعلقة بأسلوب التصفح، وهي:

أولاً: مدى التفاوت والاختلاف في اتباع أسلوب التصفح بين نوعي قواعد البيانات النصية والوسائط المتعددة ؛ حيث تعتمد أغلب قواعد البيانات النصية على تقسيم موضوعي متمثل في قوائم المصطلحات أو تقسيم عام للقطاعات الموضوعية التي تقدمها، خلاف ما تتبعه قواعد بيانات الوسائط المتعددة في الاعتماد التام على أسلوب البحث الحر بالكلمات المفتاحية دون البحث بالتصفح. وهذا يرجع إلى أن قواعد البيانات النصية غالبًا ما تقدم المحتوى النصي لموضوع أو مجال موضوعي محدد، ومن ثم كانت الحاجة لتوضيح التخصصات الفرعية والدقيقة لهذا الموضوع.

ثانيًا: تتفاوت درجة الاهتمام بالتقسيم الموضوعي بين قواعد البيانات تبعًا لعوامل مثل جودة المصادر العلمية التي تقدمها قاعدة البيانات، بالإضافة إلى طبيعة المجال ذاته. فتقدم قاعدة البيانات Eric مكنزًا يحوي تقسيمًا موضوعيًا متخصصًا في مصادر التعليم. بينما تقدم قاعدة amazon تقسيمًا عامًا لمجموعات السلع المتاحة للبيع من خلالها، وصولاً إلى قاعدة jobweb التي تكتفي بتقديم البحث الحر السريع عن الوظائف المتاحة دون الاهتمام بالتقسيم الموضوعي أو تخصص الوظائف المتاحة.

ثالثًا: يكاد ينعدم اعتماد قواعد بيانات الوسائط المتعددة على أسلوب التصفح لاسترجاع مجموعاتها، وهذا يرجع إلى طبيعة المصادر التي تقدمها هذه القواعد من صوت وفيديو يخدم أكثر ما يخدم الموسيقى والغناء والأفلام السينمائية، بما يصعب معه حصر الموضوعات المتفرقة التي تناقشها هذه الأعمال والاكتفاء بالبحث عن عناوينها أو أبطالها باستخدام البحث الحر بالكلمات المفتاحية.

رابعًا: يساهم أسلوب التصفح في التعرف إلى الكلمات الدالة وأهم الموضوعات التي تقدمها قاعدة البيانات، وهو ذات أهمية لمحركات البحث في مرحلة تكشيف قواعد البيانات، غير أن الاعتماد على أسلوب التصفح لا يؤثر في التفاعل بين محركات البحث وقواعد البيانات بالقدر الذي يوجد عليه أسلوب البحث الحر؛ حيث يقدم البحث الحر نموذجًا للبحث يتعامل به محرك البحث مع قواعد البيانات لاستدعاء تسجيلاتها الداخلية.

٢/٢/١/٤ البحث الحر:

يعد البحث الحر مفتاح الوصول السريع لمحتوى قواعد البيانات؛ حيث يمكن للمستفيد استخدام مجموعة متكاملة من حقول وآليات البحث للتعبير عن

الحاجات الموضوعية المراد الحصول عليها من قاعدة البيانات. وعلى ذلك فإن أساليب البحث الحر داخل قواعد البيانات تبلغ درجة عالية من الأهمية في استرجاع وحدات قواعد البيانات، وهذه ليست الأهمية الوحيدة لأساليب البحث الحر، وإنما يمكن الاعتماد عليها كما سيُذكر في نجاح التفاعل بين نماذج بحث قواعد البيانات وبين برامج الزاحف عند وصولها إلى واجهات بحث قواعد البيانات وتحليلها داخل محركات البحث المقترحة. وسيتم تناول أسلوب البحث الحر داخل عينة قواعد البيانات اعتمادًا على جانبين اثنين، هما:

أولاً: آليات البحث:

تمثل آليات البحث أساس عمل وبناء نموذج البحث، وذلك لكونها تحدد بدقة جوانب الربط والجمع بين كلمات البحث المختلفة وبناء استراتيجية البحث ؛ حيث يمكن لآليات البحث تغيير شكل استراتيجيات البحث بتوسيع الموضوع أو تضييق حدوده أو استثناء جانب من جوانبه... وغيرها من التغييرات التي يمكن إحداثها بين الكلمات المفتاحية لبحث قواعد البيانات. وتتنوع آليات البحث بين تلك المستخدمة مع كلمات متعددة، مثل المنطق البوليني والبحث بالتطابق، وتلك المستخدمة مع الكلمة المفردة مثل البحث بحساسية الحروف أو البحث الحر wild card. غير أن توفير آليات بحث متقدمة داخل قاعدة البيانات هو موضع اختلاف بين قاعدة وأخرى ؛ وذلك حسب أمور عدة منها تعقد المجال الموضوعي الذي تتبناه قاعدة البيانات، أو تعدد حقول الوصف الببليوجرافية داخل تسجيلات قاعدة البيانات، أو تتعدد الاتجاهات الموضوعية التي يمكن البحث بها داخل قاعدة البيانات. ويوضح الجدولان (٤-٤ / أ، ب) مجموعة من آليات البحث الأكثر شيوعًا في بحث قواعد البيانات.

الجدول (٤-٤ / أ) آليات البحث في قواعد البيانات النصية

(آليات البحث)	قواعد نصية متخصصة				
	السيرة	التربية	الحيوان	الطب	التسوق
	biography	**eric**	**Animal search**	**medlineplus**	**amazon**
البحث بالتطابق	√	√	- - -	√	√
البحث الحر ? , *	في آخر الكلمة? و❖ نصف	؟ في آخر ❖ في وسط	- - -	❖ في نهاية	√
حساسية الحروف	---	- - -	- - -	- - -	- - -
البحث بالعبارات والتوقف	---	- - -	- - - -	- - -	√
معاملات المنطق البوليني	- - -	√	And فقط	√	√
البحث برابط التسجيلة	- - -	- - -	- - -	- - -	- - -

الجدول (٤-٤ / ب) آليات البحث في قواعد بيانات الوسائط المتعددة

(آليات البحث)	قواعد وسائط متعددة					
	الصوت		صور ثابتة		صور متحركة	
	allmusic	**rhapsody**	**corbis**	**freefoto**	**movietome**	**world-of-movies**
البحث بالتطابق	√	√	√	√	- - -	- - -
البحث الحر? , *	---	√	*	❖ ؟	- - -	- - -
حساسية الحروف	- - -	- - -	- - -	- - -	- - -	- - -
البحث بالعبارات والتوقف	---	- - -	√	- - -	- - -	- - -
معاملات المنطق البوليني	---	- - - -	√	- - -		---
البحث برابط التسجيلة	---	- - -	√	- -	- - -	---

يوضح الجدولان (٤-٤ / أ ، ب) مجموعة من الملاحظات هي :

أولاً: جاء استخدام آليات البحث أكثر في قواعد البيانات النصية دون قواعد بيانات الوسائط المتعددة، ويرجع ذلك إلى أن القواعد النصية تقترب أكثر من المجالات العلمية المتخصصة، خلاف ما تكون عليه قواعد الوسائط المتعددة من عمومية المجال الموضوعي. أيضًا فإن القواعد غير المتخصصة موضوعيًا، سواء أكانت نصية أم وسائط متعددة غالبًا ما تلجأ إلى الاعتماد على الشكل البسيط من آليات البحث بالاعتماد على توفير فراغ لكتابة كلمات البحث فقط، أما تلك التي تقترب من المجالات العلمية المتخصصة فإنها تتوقع أن يكون المستفيد على دراية بآليات الربط الرياضي بين الكلمات مثل آليات الربط البوليني.

ثانيًا: جاءت آليتا البحث بالعبارات والتوقف، والبحث بحساسية الحروف كأقل آليات البحث استخداما في بحث قواعد بيانات العينة ؛ لأن قواعد البيانات هذه تتوقع من مستخدميها الإحاطة التامة بما يريدونه من موضوعات، ومن ثم فإنه لا حاجة لاستخدام البحث بحساسية الحروف طالما أن المستفيد سيكون دقيقًا في كتابة كلمات البحث، إلى جانب عدم الحاجة لآلية البحث بعبارات التوقف؛ لأن المستفيد سيكتب فقط الكلمات الدالة على الحاجة الموضوعية المخصصة، غير مضطر إلى كتابة تلك الكلمات التي لا تحمل معنى. وهو ما يمثل واقع عمل نوعي قواعد البيانات النصية المتخصصة وقواعد الوسائط العامة.

ثالثًا: اعتمدت عينة قواعد البيانات النصية والوسائط المتعددة على استخدام البحث بالتطابق بشكل أكبر من باقي آليات البحث، وذلك لكونها تفترض في مستخدم قاعدة البيانات الدقة والاحتياج إلى كل كلمة مفتاحية يضعها في فراغ البحث.

رابعًا: أثرت طبيعة بناء قواعد البيانات على درجة اهتمامها باستخدام آليات البحث ؛ حيث جاءت القواعد ذات الطابع التجاري كأكثر القواعد احتواءً على

آليات البحث ؛ مثل قاعدتي amazon في القواعد النصية، وقاعدة corbis في قواعد الوسائط المتعددة. ثم جاءت قاعدة eric المتخصصة مجال التربية كحلقة وسطى بين القواعد الترفيهية والقواعد التجارية. وهذا يعني أننا كلما ابتعدنا عن المجالات العلمية المتخصصة أو الطبيعة التجارية الخالصة واقتربنا إلى الجوانب الترفيهية، فإننا نفقد الاهتمام باستخدام آليات البحث شيئًا فشيئًا.

خامسًا: افتقرت قاعدتا movietome, world-of-movies إلى أية آليات بحث يمكن استخدامها لاسترجاع محتوى قاعدة البيانات ؛ وذلك يرجع إلى أن هذه القواعد ترفيهية بالدرجة الأولى، وتعتمد بشكل رئيس على مجرد البحث بالكلمات المفتاحية في فراغ واجهة البحث، ثم استدعاء أي من النتائج التي قد تبعد أو تقترب من كلمات البحث.

ثانيًا: حقول البحث:

يعتمد البحث بالحقول في أي قاعدة بيانات على مدى تبني هذه القاعدة لبناء تسجيلات ببليوجرافية لمجموعاتها، وذلك حتى يمكن لها استرجاع هذه التسجيلات باستخدام البحث بحقول الوصف الببليوجرافي. وتمثل حقول الوصف الببليوجرافي في قواعد البيانات عنصرا أساسيا من عناصر الاسترجاع ؛ حيث تتكون قاعدة البيانات من جداول مختلفة ممثلة في شكل سجلات أو تسجيلات تحوي بداخلها البيانات المنظمة في شكل حقول. وكما سبق الذكر في الفصول السابقة فإن الحقول التي تمثل أعمدة الجداول هي أصغر وحدة بناء في جسم قاعدة البيانات. وعلى الرغم من الأهمية الكبيرة لحقول الوصف الببليوجرافي، فقد لاحظ الباحث أن عينة الدراسة من قواعد البيانات قد أغفلت البحث بحقول الوصف الببليوجرافي. وهذا يمكن إرجاعه إلى أن قواعد بيانات الويب، سواء تلك النصية المتخصصة أو التي تحتوي على وسائط متعددة عامة ما زالت تتعامل مع المستفيد ذات القدرات المحدودة في البحث، معتمدة في ذلك على توفير فراغ للبحث توضع به كلمات البحث مباشرة دون تخصيص عناصر بحث أخرى

تبحث داخل حقول التسجيلات. ويوضح الجدولان (٤-٥/ أ ، ب) واقع البحث بالحقول داخل عينة قواعد البيانات.

الجدول (٤- ٥ / أ) حقول البحث داخل قواعد البيانات النصية(*)

نموذج البحث (حقول البحث)	قواعد نصية متخصصة				
	السيرة	التربية	الحيوان	الطب	التسوق
	biography	eric	Animal search	medlineplus	amazon
المؤلف أو المنشئ	- - -	√	- - -	- - -	- - -
العنوان	- - -	√	- - -	- - -	- - -
رؤوس الموضوعات	- - -	√	- - -	- - -	- - -
الجهة أو الناشر	- - -	√	- - -	- - -	- - -
النوع أو الشكل	- - -	√	- - -	- - -	√
رابط أو محدد التسجيلة ID	- - -	- - -	- - -	- - -	√
تواريخ مرتبطة	- - -	√	- - -	- - -	- - -

الجدول (٤-٥ / ب) حقول البحث داخل قواعد بيانات الوسائط المتعددة

(حقول البحث)	قواعد وسائط متعددة					
	الصوت		صور ثابتة		صور متحركة	
	allmusic	rhapsody	corbis	freefoto	movietome	world-of-movies
المؤلف أو المعد	√	√	√	---	---	---
العنوان	√	√	√	---	---	---
رؤوس الموضوعات	√	√	√	---	---	---

(*) اعتمد الباحث في اختيار الحقول الواردة في الجداول تحديدًا على عناصر الوصف الواردة في معيار dublin core، وذلك لكونه معيارًا يحمل عناصر وصف عامة تلائم حقول الوصف الببليوجرافي الواردة في قواعد البيانات النصية المتخصصة أو قواعد بيانات الوسائط المتعددة العامة.

الجهة أو الناشر	---	---	---	---	---	---
النوع أو الشكل	√	√	√	---	---	---
رابط أو محدد التسجيلة ID	---	---	√	---	---	---
تواريخ مرتبطة	---	---	√	---	---	---

يؤكد الجدولان السابقان (٤-٥ / أ ، ب) على الجوانب التالية:

أولاً: ثمة تطابق يكاد يوجد بين جدولي آليات البحث وجدولي حقول البحث. حيث يمكن القول أن درجة الاهتمام التي وفرتها كل قاعدة بيانات من عينة الدراسة لآليات البحث هي ذاتها التي قد وفرتها للاهتمام بحقول البحث، وقد خالف ذلك فقط قاعدة البيانات amazon التي أولت اهتمامًا كبيرًا لآليات البحث، ولم تهتم بالقدر نفسه لحقول البحث. ثم يأتي التقارب بين القواعد الباقية في آليات وحقول البحث، وهو ما وضح جليًا مع قاعدتي بيانات الأفلام movietome, world-of-movies.

ثانيًا: حققت قاعدتا eric في التربية وقاعدة corbis في الصور أكبر درجة اهتمام بتوفير حقول البحث عن مجموعاتها، وهذا بدوره يعكس البناء الداخلي الجيد للتسجيلات الببليوجرافية داخل هذه القواعد، واعتمادها على حقول وصف دقيقة تصف المجموعات الداخلية. ويرجع الباحث ذلك إلى التخصص العلمي لقاعدة بيانات eric والنشاط التجاري لقاعدة بيانات corbis اللذين دفعا قاعدتي البيانات إلى توفير تسجيلات وصف للمجموعات، ثم تنشيط حقول الوصف هذه حتى يتمكن مستخدمو قاعدة البيانات من إخضاع حقول الوصف للبحث واستدعاء المجموعات.

ثالثًا: تعكس حقول البحث التي توجد في واجهات بحث قواعد البيانات مدى تعقد أو بساطة نماذج البحث داخل قواعد البيانات؛ فكلما زادت كثافة حقول البحث، كان نموذج البحث معقدًا، والعكس صحيح. وهذا بدوره يعكس حاجة نماذج البحث المعقدة إلى قدر كبير من التحليل واللجوء إلى المعيارية في العمل داخل قواعد البيانات.

رابعًا: أن الجانب الأكثر أهمية من الجوانب السابقة هو أن ثمة علاقة بين نموذج البحث في واجهة بحث قاعدة البيانات، الذي يتكون من حقول بحث، وبين برنامج الزاحف الذي سيقوم بالدخول إلى قاعدة البيانات باستخدام بوابة نموذج البحث. وبشكل مباشر فإن برنامج الزاحف سوف يستخدم نموذج واجهة بحث قاعدة البيانات في ضمها وإضافتها إلى محرك بحث قواعد البيانات المقترح في هذه الدراسة.

خامسًا: تأخذ حقول البحث جانبًا آخر من الأهمية في نموذج بحث واجهات قواعد البيانات، هو أنها سوف تستقبل فيما بعد من محرك البحث مجموعات استفسارات المستفيدين ومعالجتها ثم الرد عليها، ثم توجيه مجموعة النتائج المستدعاة مرة أخرى إلى محرك البحث. وعلى ذلك فإن البساطة والاقتراب من المعيارية هما سبيلا حقول بحث قاعدة البيانات لتحقيق أفضل درجات العمل والأداء.

إن برامج الزاحف داخل محركات البحث هي المنوطة بأداء عمليات الإضافة لقواعد البيانات الجديدة، وعلى ذلك فإن قدرة برامج الزاحف على تحليل وفك تركيبة نماذج واجهات البحث في قواعد البيانات هي المعيار الرئيسي في نجاح محرك بحث قواعد البيانات لضم كل قواعد بيانات الويب غير المرئية التي يستطيع الوصول إليها. لقد عملت برامج الزاحف داخل محركات البحث التقليدية على تحليل الروابط الفائقة داخل مواقع الويب، أما عملها مع قواعد بيانات الويب هنا فيعتمد بشكل أساسي على تحليل وفهم نموذج بحث قاعدة البيانات بعناصره المختلفة من آليات وحقول وأوامر واجهات قواعد البيانات.

٣/١/٤ مستويات الوصول:

تحمل قواعد بيانات الويب محتوى غالبًا ما يختلف في طبيعته عن صفحات ومواقع الويب الأخرى، وذلك لما لهذا المحتوى من خصائص تتنوع بين دقة الصياغة

والبناء وشدة التقييد في الضبط والتحكم؛ حيث تبتعد قواعد البيانات أكثر ما تبتعد عن المحتوى الضعيف الشائع في مواقع الويب الأخرى. وعلى ذلك فإن المحتوى الذي تقدمه هذه القواعد دائمًا ما يشترط للوصول إليه إما المقابل المادي أو الاشتراك أو الانتماء إلى هيئات علمية يمكن للمستخدم من خلالها الحصول على محتوى قواعد البيانات. وتعد هذه الشروط بمثابة درجات تختلف فيها قواعد البيانات ذات النشاط التجاري عن قواعد البيانات العلمية المتخصصة إلى قواعد البيانات الترفيهية التي قد لا تفرض أيَّ شروط للوصول إلى محتوياتها.

وتمثل قضية مستوى الوصول إلى محتوى قواعد بيانات الويب حجر عثرة أمام محاولات تكشيف هذه القواعد وإتاحتها، ذلك لأن الكثير من قواعد البيانات لا يمكن النفاذ إلى محتواها دون الرسوم المادية، وهو ما يمكن تناوله أثناء الحديث عن القضايا الجانبية المرتبطة بتكشيف قواعد بيانات الويب غير المرئية. إلا أن الباحث يقترح بشكل مبدئي صياغة نموذج من اتفاقيات أو إجازات الوصول licenses إلى المحتوى يتم إبرامها بين محركات البحث وبين قواعد البيانات، يحمل من خلالها برنامج الزاحف إجازات الدخول لمحتوى قواعد البيانات التجارية. ويوضح الجدولان (٤-٦ / أ، ب) واقع تحديد الوصول وضبط المحتوى داخل قواعد بيانات العينة.

الجدول (٤- ٦ / أ) مستويات الوصول داخل قواعد البيانات النصية

مستويات الوصول (حدود الإتاحة)		قواعد نصية متخصصة				
		السيرة	التربية	الحيوان	الطب	التسوق
		biography	eric	Animal search	medlineplus	amazon
مستوى الوصول	محدد	√	- - -	- - -	- - -	√
	مفتوح	- - -	√	√	√	- - -
المقابل	بمقابل	√	- - -	- - -	- - -	√
	مجانية	- - -	√	√	√	- - -
الاشتراك	هيئات	- - -	- - -	- - -	- - -	- - -
	أفراد	√	- - -	- - -	- - -	√

الجدول (٤– ٦ / ب) مستويات الوصول داخل قواعد بيانات الوسائط المتعددة

مستويات الوصول (حدود الإتاحة)		قواعد وسائط متعددة					
		الصوت		صور ثابتة		صور متحركة	
		allmusic	rhapsody	corbis	freefoto	movietome	world-of-movies
مستوى الوصول	محدد	√	داخل أمريكا√	√	---	√	√
	مفتوح	---	---	---	√	---	---
المقابل	بمقابل	√	---	√	---	√	√
	مجانية		√	---	√	---	---
الاشتراك	هيئات	---	---	---	---	---	---
	أفراد	√	√	√	---	√	√

يعكس الجدولان (٤-٦ / أ ،ب) مجموعة من الملاحظات يمكن تناولها من خلال النقاط التالية:

١/٣/١/٤ مستوى محدد أو مفتوح:

أولاً: وضعت قواعد بيانات العينة درجتين من درجات الحصول على محتواها الداخلي ؛ هما المستوى المفتوح، وذلك حيث يمكن للمستفيد الوصول إلى محتوى التسجيلات الداخلية لقاعدة البيانات والاطلاع على الملف الأصلي، سواء كان النص الكامل أو ملف وسائط متعددة، ومستوى آخر هو المستوى المحدد الذي يقف عند تمكين المستفيد من إجراء البحث والوصول إلى مكونات التسجيلة النهائية، غير أنه لا يمكن الاطلاع على ملفات النصوص أو الوسائط المتعددة التي توجد داخل تسجيلات قواعد البيانات.

ثانيًا: بدا واضحًا أن المستوى المحدد إنما يرتبط بالصبغة التجارية لقاعدة البيانات، وهو ما قد يتدرج من قواعد البيانات العلمية المتخصصة eric, medline

مـرورًا بقاعـدة البيانـات التجاريـة amazon وصـولاً إلى قواعـد بيانـات الوسـائط المتعددة التجارية التي تحمل موادَّ ترفيهية تستقطب الكثير من مستخدمي الويب الراغبين في الحصول على هذه المواد.

ثالثًا: ظهور نمط مختلف من مستوى الوصول إلى المحتوى الداخلي في قاعدة بيانات الموسيقى rhapsody ؛ حيث ربطت هذه القاعدة إمكانية الوصول وفتح الملفات في تسجيلاتها الداخلية بالحدود المكانية ؛ فمن كان من المستفيدين من خارج الولايات المتحدة الأمريكية فليس له الحق في الوصول إلى ملفات الموسيقى داخل قاعدة البيانات، على الرغم من كونها قاعدة مجانية.

٢/٣/١/٤ خدمة مدفوعة أو مجانية:

تستمد قواعد بيانات الويب طبيعتها التجارية أو المجانية من الهيئات التي تتولى إنشاء هذه القواعد ؛ ويعكس ذلك قاعدتا البيانات eric, medline اللتان تقف خلف بنائهما هيئتان علميتان متخصصتان، إحداهما في مجال التربية والتعليم، والأخرى في مجال الطب والصحة. ومن ثم كانت إتاحة محتوى هاتين القاعدتين من قبيل تحقيق الإفادة العلمية للباحثين والمتخصصين في مجال التربية والطب فهما هيئتان علميتان تخدمان البحث العلمي ولا تبغيان الربح. وهذا خلاف قواعد البيانات الأخرى التي يقف خلف بنائها إما هيئات أو أفراد ذوو نوايا تجارية.

٣/٣/١/٤ اشتراك الهيئات أو الأفراد:

إن مصطلح الاشتراك أو الانضمام لقاعدة البيانات هنا يحمل معنيين مختلفين ؛ أولهما : الاشتراك في قاعدة البيانات بدفع رسوم تتيح للمستخدم الوصول إلى المحتوى الداخلي لقاعدة البيانات لمدة زمنية محددة، وثانيهما : الاشتراك في قاعدة البيانات والحصول على اسم مستخدم وكلمة مرور user ID, Password دون أن يتكلف المستخدم دفع رسوم للحصول على المحتوى، وهي الحالة التي تحمل فيها قاعدة البيانات صفة المواقع المعلنة، ومن ثم فإن الصبغة التجارية لها تكمن في رسوم الإعلانات، لا في رسوم الاشتراكات من المستخدمين.

ثمة ارتباط واضح بين جانب الاشتراك وبين الخدمات مدفوعة الأجر في قواعد بيانات العينة ؛ حيث يمكن القول إن الاشتراكات هي المنفذ الأول لدفع الرسوم المقررة من قبل قاعدة البيانات. وقد بدا واضحًا، خاصة في قواعد الوسائط المتعددة، أن الاشتراك يركز أكثر على الأفراد، فمن البديهي أن يسعى مستخدمو الويب ذوو الطبيعة الفردية الشخصية للوصول إلى المواد الترفيهية. ويعكس ذلك واقع الأمر في قاعدتي الأفلام movietome, world-of-movies، ثم بدأ ينحسر وصولاً إلى قواعد البيانات العلمية المتخصصة eric, medline.

٤/١/٤ خدمات قواعد البيانات:

تمثل خدمات قواعد البيانات الجانب الثاني الأهم بعد آليات وأساليب البحث في عمل قواعد بيانات الويب. وتختلف قواعد بيانات الويب عن غيرها من قواعد البيانات ؛ وذلك لما تتطلبه قواعد الويب من تخصيص واجهات عمل وخدمات وخصائص تمكنها من أداء مهامها في بيئة تمتاز بعنصري التغير والديناميكية المستمر. إن مستخدمي الويب غالبًا ما يسعون إلى مواقع الويب التي تقدم خدمات أكثر من غيرها من المواقع في المجال أو الاهتمام نفسه؛ حيث يحتاج بعض المستخدمين إلى قدرات التشغيل العالية، ويحتاج بعضها الآخر إلى إمكانية الحفظ والتحميل downloading للمعلومات، كما يسعى بعضها الآخر إلى المواقع التي توفر أساليب بحث واسترجاع متقدمة تمكنه من الوصول السريع إلى المعلومات المرجوة، هذا فضلاً عن جوانب الشمولية والتغطية الواسعة لمعلومات المجال المتخصص.

وقد قامت الدراسة هنا بالتركيز على أهم الخدمات التي تحوز اهتمام مستخدمي الويب، التي تلائم عمل قواعد البيانات بعينها دون غيرها من مواقع الويب الأخرى. وجدير بالذكر أن ثمة فارق كبير بين الخدمات التي تقدمها قواعد البيانات العلمية المتخصصة وبين قواعد البيانات الترفيهية أو العامة ؛ ذلك لأن احتياجات الباحثين في قواعد البيانات العلمية يمكن أن تنحصر في توفير النصوص الكاملة أو توفير أساليب البحث المتقدمة، أما احتياجات مستخدمي

قواعد البيانات الترفيهية العامة فيمكن أن تنحصر في قدرات تحميل المعلومات وحفظها، وتوفير التنوع والتعدد في المواد الصوتية والمرئية.

يوضح الجدولان (٤-٧ / أ ، ب) أهم الخدمات التي يمكن أن تقدمها قواعد بيانات الويب، وهي :

الجدول (٤- ٧ / أ) خدمات قواعد البيانات النصية

خدمات قواعد البيانات (الخدمات المصاحبة)	قواعد نصية متخصصة				
	السيرة	التربية	الحيوان	الطب	التسوق
	biography	eric	Animal search	medlineplus	amazon
تأريخ عمليات البحث	- - -	√	- - -	- - -	- - -
القفز بين الصفحات	√	√	√	√	√
التحميل للملفات المنتقاة	- - -	√	- - -	- - -	- - -
حفظ الصفحات وطباعتها	√	√	√	√	√
التواصل مع فريق العمل	√	√	- - -	√	- - -
توفير مواد ذات كثافة عالية	√	- - -	√	√	√
تقارب فترات التحديث	√	√	٢٠٠٣	√	√

الجدول (٤- ٧ / ب) خدمات قواعد بيانات الوسائط المتعددة

خدمات قواعد البيانات (الخدمات المصاحبة)	قواعد وسائط متعددة					
	الصوت		صور ثابتة		صور متحركة	
	allmusic	rhapsody	corbis	freefoto	movietome	world-of-movies
تأريخ عمليات البحث	---	√	---	√	---	---
القفز بين الصفحات	√	√	√	---	√	√
التحميل للملفات المنتقاة	---	داخل أمريكا	---	√	---	---
حفظ الصفحات وطباعتها	√	√	√	√	√	√

التواصل مع فريق العمل	√	- - -	√	√	- - -	√
توفير مواد ذات كثافة عالية	√	√	√	√	- - -	√
تقارب فترات التحديث	√	√	√	√	√	√

يعكس الجدولان (٤-٧ / أ، ب) بعض الملاحظات يتم تناولها في العناصر التالية :

١/٤/١/٤ خدمة تأريخ عمليات البحث:

اهتم القليل من قواعد البيانات بتقديم تأريخ لعمليات البحث التي يقوم بها مستخدمو قاعدة البيانات، وهو ما قد يكون ذا أهمية كبيرة عند الكثير من المستفيدين. وتكمن أهمية تأريخ عمليات البحث في كونه يضع قائمة بمختلف المصطلحات البحثية والكلمات المفتاحية التي استخدمها المستفيد في بحث قاعدة البيانات، حتى يمكن من خلال النظر فيها ومراجعتها أن يصحح ويغير المستفيد من مسار بحثه، فقد تشير هذه القائمة للمستفيد بإضافة مصطلحات أخرى أو حذف بعض من كلمات البحث، أو الاستعانة بالمترادفات ... وغيرها من الإشارات التي يمكن استنتاجها من قائمة المصطلحات المستخدمة في البحث. ويمكن تمثيل خدمة تأريخ البحث هنا بمثابة المراجع والمرشد للمستفيد في تحديد مسارات البحث. وكانت أفضل قواعد البيانات استخدامًا لهذه الخدمة هي قاعدة بيانات التربية eric، وهو ما يتماشى مع قدرات هذه القاعدة في البحث وتقديم المعلومات المحكمة المتخصصة.

٢/٤/١/٤ إمكانات التنقل السريع بين الصفحات:

توفر خدمة القفز بين الصفحات قدرة المستفيد على الانتقال السريع من موقع إلى آخر داخل قاعدة البيانات، ومن مسار بحث إلى مسار آخر. وتتمثل هذه الخدمة في ما يسمى بخريطة الموقع site map، حيث تجمع هذه الخريطة مختلف

الروابط وعناوين الصفحات داخل موقع قاعدة البيانات بما يتيح للمستفيد تحديد عناوين في مستويات وصول أعمق من تلك الصفحات المرتبطة بواجهة البحث بشكل مباشر. وتكاد تتفق قواعد بيانات عينة الدراسة في هذه الخدمة؛ لسهولة تقديمها، فضلاً عن كونها أحد المكونات الأساسية لواجهات عمل الكثير من مواقع الويب.

٣/٤/١/٤ خدمات تحميل الملفات المنتقاة:

ترتبط خدمة تحميل الملفات غالبًا بطبيعة مجموعات قاعدة البيانات؛ حيث ترتبط كلمة تحميل أو downloading بملفات الصوت والصورة، فضلاً عن ملفات النصوص الكاملة في شكل doc, pdf. وقد عكست الدراسة التجريبية الارتباط الوثيق بين خدمات تحميل الملفات وبين الطبيعة التجارية التي تحكم قواعد البيانات، وقد تجلى ذلك في قواعد بيانات الوسائط المتعددة التي تحمل قدرًا هائلاً من ملفات الموسيقى والصور والأفلام، وهذه المواد تمنح فقط عن طريق دفع المقابل المادي للحصول عليها. وتمثل ذلك في قواعد البيانات corbis, movietome, world of movies, all music. وقد خالف هذا الأمر قاعدة النصوص الكاملة التربوية eric في كونها تتيح الوصول المجاني إلى ملفات المقالات الكاملة.

٤/٤/١/٤ إمكانات حفظ الصفحات وطباعتها:

تعمل بعض قواعد بيانات الويب على إتاحة القراءة فقط لمجموعاتها دون إتاحة خاصية الحفظ أو الطباعة للمستفيد، وذلك غالبًا ما يكون في قواعد البيانات النصية التي تتيح محتوياتها للقراءة فقط. ويكون حجب الحفظ هنا في شكل رسالة توجهها قاعدة البيانات إلى المستفيد في حالة إعطاء أمر الحفظ أو الطباعة لحاسبه الشخصي PC. وقد اتفقت جميع قواعد بيانات عينة الدراسة على إتاحة هذه الخاصية أمام المستفيد، ويرجع الباحث ذلك إلى أن قواعد بيانات الويب الآن تمتلئ بكثير من أشكال الملفات سواء النصية أو المصورة أو الصوتية التي لا يتم

عرضها في واجهات قاعدة البيانات، وإنما تحتاج إلى برامج خاصة لعرضها، وعلى ذلك فليس هناك حاجة لغلق أمر الطباعة والحفظ في واجهاتها، طالما أنه يمكن التحكم بملفاتها الداخلية.

٥/٤/١/٤ خدمات التواصل مع فريق العمل:

تعد سمة التفاعلية الملمح الأساسي في عمل مواقع الويب، ومن ثم فإن قواعد البيانات تحتاج إلى دعم هذه الخاصية حتى تحقق جانب التواصل مع مستخدميها. وتقوم خاصية التواصل مع فريق العمل هنا بدور التغذية المرتدة feedback في النظم المفتوحة مثل قواعد بيانات الويب، ومن ثم يمكن التعرف إلى آراء واتجاهات مستخدمي قواعد البيانات. كما أن عنصر التغذية المرتدة يمكّن قاعدة البيانات من تصحيح مسار العمل إذا ما كانت هناك ملاحظات من جانب المستفيدين. وقدمت قاعدة البيانات eric أفضل شكل لخدمة التواصل بين كل قواعد عينة الدراسة، حيث توفر لهذه الخدمة صفحة خاصة يمكن للمستفيد ملء النموذج والإرسال المباشر إلى فريق عمل قاعدة البيانات، أما باقي القواعد الأخرى فقد قدمت خدمة التواصل في شكل بيانات الاتصال الممثلة في البريد الإلكتروني وبيانات الموقع إذا ما وجدت.

٦/٤/١/٤ توفير مواد ذات كثافة عالية:

تتقدم بعض قواعد البيانات بتوفير أنواع محددة من التقنيات إذا ما تخصصت في شكل أو نوع بعينه من وسائط حمل المعلومات ؛ فتحتاج قواعد بيانات المواد المصورة إلى مواد مصورة ذات كثافة عالية high resolution حتى تستقطب إليها مستخدمي الويب الباحثين عن هذه المواد. ويمكن القول إن هذا الشأن يقتصر على قواعد بيانات الوسائط المتعددة دون قواعد البيانات النصية. وتتفق جميع مواقع الويب في الحاجة إلى عرض موادها بأعلى إمكانية أمام المستفيدين، وقد لاحظ الباحث ذلك في قواعد بيانات الصور والأفلام، غير أن الأمر قد شكل

أحد جوانب العمل بين قاعدة البيانات وبين المستفيد ؛ كما في قاعدة بيانات corbis للصور التي تتيح تحميل الصورة بدرجات كثافة متفاوتة، وبقدر ما تكون درجة كثافة الصورة عالية يكون المقابل المادي أعلى.

٧/٤/١/٤ تقارب فترات التحديث:

إن الويب عامة، وقواعد البيانات خاصة، يجب أن تتسم بجانب التحديث المستمر للمجموعات، لأن المستفيد ما يزال يقبل على قاعدة البيانات التي تهتم بالتحديث المستمر لموادها وبياناتها. وان كان ذلك يعد أمرًا طبيعيًا في مواقع الويب، إلا أن الباحث يرى أن عامل التحديث بالنسبة لقواعد البيانات هو أهم المعايير التي يبنى على أثرها قرار التعامل مع قاعدة بيانات دون الأخرى. وقد شذت قاعدة بيانات animal search عن باقي قواعد بيانات عينة الدراسة من حيث الالتزام بجانب التحديث ؛ حيث توقف تحديث المجموعات داخلها عند عام ٢٠٠٣م.

إن القرار الذي تتخذه قاعدة البيانات بشان التحديث سوف يؤثر فيما بعد على التفاعل بينها وبين برنامج الزاحف ؛ حيث سيقوم برنامج الزاحف بزيارات دورية إلى قاعدة البيانات للوقوف على الموضوعات أو المجموعات حديثة الإضافة لقاعدة البيانات، وذلك من خلال طرح الأسئلة وتلقي الإجابات وتحليلها.

الخصائص النوعية لقواعد البيانات :

تختلف قواعد البيانات فيما بينها في كثير من خصائص العمل، وهذا يرجع في أغلب الأحيان إلى طبيعة المحتوى الداخلي لقاعدة البيانات وشكل المعلومات التي تحتويها. وعلى ذلك فإن نوعي قواعد بيانات عينة الدراسة النصية والوسائط المتعددة يمتلك كل منها ملامح للعمل تلائم طبيعة المواد النصية أو خيارات للتعامل مع المواد المصورة أو المسموعة في قواعد الوسائط المتعددة. بالإضافة إلى شكل البيانات فإن طبيعة قاعدة البيانات التجارية المرسمة أو العلمية المجانية من

شـأنها أن تفـرض ملامـح محـددة للعمـل داخـل قاعـدة البيانـات ؛ حيـث تـزدحم واجهـات قواعـد البيانـات التجاريـة المتخصـصة في الوسـائط المتعـددة بكـثير مـن اللقطات والإعلانات عـن مجموعاتها الداخلية إلى جانب تـداخل الـصور الثابتة أو المتحركة، وهو ما يختلف بـدوره تمامًا عـن قواعـد البيانات النـصية العلميـة الـتي تمتاز بواجهات بحث بسيطة تكاد تخلو من الإعلانات أو الصور.

إن الخصائص النوعية لقاعدة البيانات لا تؤثر فقط في شكل واجهات التعامل مـع قاعـدة البيانـات، وإنمـا تـؤثر أيـضًا في أسـاليب العمـل والبحـث داخـل قاعـدة البيانـات. فتعمـل قواعـد بيانـات الوسـائط المتعـددة علـى الاسـتعانة بأسـاليب تقنيـة للبحث تخرج عن أساليب البحث بالكلمات المفتاحيـة مثل الاسـترجاع المبني علـى المحتـوى content based retrieval CBR الـذي لا يـتم البحـث داخلـه بالكلمـات المفتاحية وإنما يعتمد على بنيـة المواد المصورة ولونها وشكلها. ومـن هـذا المنطلـق آثـر الباحـث الحـديث عـن واقـع الخـصائص النوعيـة لقواعـد البيانـات بـشكل منفصل، وذلك في إطار العناصر التالية :

٥/١/٤ أنماط عرض نتائج البحث:

تحتل خدمـة عـرض النتـائج مكانًـا بـالغ الأهميـة بـين بـاقي جوانـب عمـل وأداء قواعد بيانات الويب. وأبرز ما يمكـن ذكـره أن نتـائج البحـث تمثـل الإجابـة أو ردة الفعل على استفسار المستفيد المصاغ في مجموعة كلمات مفتاحيـة، إلى جانـب أن نتائج البحث المسترجعة تمثل مـدى نجـاح قاعـدة البيانـات أو فشلها في التفاعـل مـع اسـتفسار المستفيدين. كما أن مقياسـي الاسـتدعاء والتحقيق يتم تفعليهمـا اعتمـادًا علـى نتـائج البحـث لتحديـد درجـة الكفـاءة في نظـم الاسـترجاع، ومنهـا قواعـد البيانات. وإذا كانت واجهة بحث قاعدة البيانات على قدر من الأهميـة في صـياغة واسـتقبال استفسـارات المـستفيدين، فـإن واجهـة عـرض النتـائج لا تقـل أهميـة عـن ذلـك في عـرض الوحـدات المـسترجعة وترتيبهـا بحـسب الـصلة بموضـوع بحـث

المستفيد. إن قواعد بيانات الويب يجب أن تفكر في جوانب عدة عند بنائها لواجهات عرض النتائج مثل ؛ طبيعة البيانات المعروضة سواء أكانت نصية أو وسائط متعددة، وأيضًا عدد الوحدات التي سيتم عرضها في كل صفحة، بالإضافة إلى مجموعة الخيارات أو البيانات المصاحبة التي ستشرح وتوضح طبيعة كل تسجيلة من التسجيلات المعروضة.

وجدير بالذكر أن الاهتمام بواجهات عرض النتائج قد يختلف من قاعدة بيانات إلى أخرى بحسب طبيعة المحتوى الداخلي لمجموعات قاعدة البيانات ؛ حيث تحتاج قواعد بيانات الوسائط المتعددة إلى الاهتمام بعوامل العرض والتنظيم والإظهار للتسجيلات المعروضة، التي ستتمثل في أغلب الأحيان في لقطات مصورة ثابتة أو متحركة. وسوف يتم الحديث عن عرض النتائج داخل قواعد بيانات العينة في جوانب مثل ؛ خصائص ترتيب النتائج المعروضة، والتحكم في طرق عرض النتائج، والوصول إلى ملفات التسجيلات، والبيانات المصاحبة للتسجيلات المعروضة.

إن محرك بحث قواعد بيانات الويب غير المرئية المقترح في هذه الدراسة سوف يهتم بجانب عرض التسجيلات وترتيبها داخل قواعد البيانات التي يتعامل معها. لأن عمل هذا المحرك سيبنى على تحويل الاستفسارات البحثية الواردة من المستفيدين إلى قاعدة البيانات أو قواعدها، ثم تلقي الإجابات في شكل قوائم بالتسجيلات الملائمة للرد على استفسار البحث، ومن ثم يقوم محرك بحث قواعد بيانات الويب غير المرئية بتجميعها وإعادة تنظيمها وعرضها مرة أخرى أمام المستفيد. وهذا العمل سيكون شبيهًا بما يقوم به محرك البحث المتعدد met search engine الذي يبحث في مجموعة محركات بحث مختلفة.

ويوضح الجدول (٤– ٨) والجدول (٤– ٩) واقع الجوانب المختلفة لعرض النتائج داخل عينة قواعد بيانات الويب غير المرئية. وهي كالتالي:

الجدول (٤- ٨) أنماط عرض نتائج البحث

أنماط عرض نتائج البحث		قواعد نصية متخصصة				
		السيرة	التربية	الحيوان	الطب	التسوق
		biography	eric	Animal search	medlineplus	amazon
خصائص الترتيب ranking	درجة الصلة	√	√	√	√	√
	حداثة المواد	---	√	---	---	√
	أنماط أخرى	---	مؤلف عنوان شكل	---	---	سعر، مراجعة
عرض النتائج المستدعاة	اختياريا	---	√	---	---	√
	تلقائيا	√	√	√	√	√
الوصول لملفات التسجيلات	كرابط أولا	---	√	---	---	√
	عرض مباشر	√	---	√	√	---
البيانات المصاحبة للنتائج	بيانات واصفة	√	√	√	مسار فقط	√
	خيارات العمل	---	√	---	---	√
عدد الوحدات المعروضة		١٠	اختياري	١٥	١٠	١٠
البحث داخل النتائج		- - -	√	- - -	- - -	- - -
عرض النتائج	روابط	√	- - -	√	√	
	تسجيلات	- - -	√	- - -	- - -	√

الجدول (٤- ٩) أنماط عرض نتائج البحث في قواعد بيانات الوسائط المتعددة

أنماط عرض نتائج البحث		قواعد وسائط متعددة					
		الصوت		صور ثابتة		صور متحركة	
		allmusic	rhapsody	corbis	freefoto	movietome	world-of-movies
خصائص الترتيب ranking	درجة الصلة	√	√	√	√	√	√
	حداثة المواد	---	---	√	---	---	---
	أنماط أخرى	---	---	---	---	---	---

عـرض النتـائج	اختياريا	---	---	√	---	---	---
المستدعاة	تلقائيا	√	√	---	√	√	√
	كرابط أولا	---	√	√	√	√	√
الوصـول لملفـات التسجيلات	عرض مباشر	√	---	√ لقطات	√ لقطات	√ لقطات	---
	بيانات واصفة	---	√	√	√	√	---
البيانـــــات المصاحبة للنتائج	خيارات العمل	---	---	√	---	---	---
عدد الوحدات المعروضة		√	√	غير محدد	---	√	غير محدد
البحث داخل النتائج		---	---	---	---	---	---
	روابط	---	---	√	√	---	√
عرض النتائج	تسجيلات	√	√	√	---	√	---

يعكس الجدولان (٤– ٨) و(٤– ٩) مجموعة من الملاحظات المهمة، هي :

١/٥/١/٤ خصائص الترتيب ranking :

أولاً: لقد اتفقت قواعد البيانات النصية وقواعد الوسائط المتعددة فيما يتبعونه من أنماط ترتيب النتائج ؛ حيث اعتمدت أغلب قواعد بيانات العينة في شقيها النصية وقواعد الوسائط المتعددة مبدأ الترتيب بحسب الصلة بالموضوع subject relevance كمبدأ أساسي لترتيب مجموعة النتائج المسترجعة من قاعدة البيانات. وقد خالف ذلك في كل من القواعد النصية وقواعد الوسائط قواعد البيانات eric, amazon, corbis في كونها قد زادت على مبدأ الترتيب بحسب الصلة أنماطًا أخرى مثل: الترتيب وفقًا للسعر في amazon، والترتيب وفقًا لتهجئة اسم المؤلف أو العنوان في eric، والترتيب وفقًا لتواريخ التقاط الصورة في corbis.

ثانيًا: إن ما يعنينا هنا هو الاهتمام بأنماط ترتيب النتائج التي تعمل بها قواعد بيانات الويب ؛ ذلك لما له من تأثير مباشر في الإبقاء على بعض النتائج واستبعاد بعضها الآخر في مرحلة التجميع والفرز التي سيعمل بها محرك بحث قواعد البيانات المقترح في هذه الدراسة. ويرى الباحث أن قواعد بيانات الويب عليها الالتزام بجعل مبدأ الترتيب بحسب الصلة المبدأ الموحد لترتيب النتائج المسترجعة، حتى يمكن العمل بيسر عند تجميع النتائج الواردة من أكثر من قاعدة بيانات وفرزها كلية ثم ترشيح مجموعة محددة للعرض أمام المستفيد على واجهة محرك بحث قواعد البيانات.

٢/٥/١/٤ اختيار طرائق العرض أو العرض التلقائي:

تعمل بعض قواعد البيانات على توفير الاختيار أمام المستفيد في عدد وحجم وطريقة عرض النتائج المسترجعة من جانب قاعدة البيانات. غير أن واقع عينة الدراسة يؤكد وجود صلة وثيقة بين اعتماد قواعد البيانات لمبدأ الترتيب بحسب الأقرب للموضوع وبين عرض النتائج ؛ حيث تعرض النتائج بشكل تلقائي مرتبة بحسب الأقرب للموضوع من أعلى إلى أسفل غير تاركة للمستفيد اختيار أي شكل آخر للترتيب. وقد انسحب ذلك على كامل عينة الدراسة خلا قواعد البيانات eric, amazon, corbis.

تختلف طرائق عرض النتائج بين قاعدة بيانات وأخرى، وذلك بحسب المحتوى الداخلي لقاعدة البيانات؛ حيث يمكن أن تكتفي قواعد البيانات النصية بمجموعة من البيانات البسيطة المحددة أو رابط فقط للتسجيلة النهائية داخل قاعدة البيانات، بينما تعمد قواعد بيانات الوسائط المتعددة إلى عرض لقطات مصغرة thambinals في حالة ملفات الصور، حتى يمكن للمستفيد التقاط فكرة أولية عن الملف النهائي للصورة، بالأسلوب نفسه تسير قواعد بيانات الأفلام في عرض لقطات ثابتة أو متحركة لمجموعة ملفاتها الداخلية.

٣/٥/١/٤ أنماط الوصول لملفات الوسائط:

إن الويب منذ أن تخطت حدود الملفات النصية إلى الأنواع الأخرى من الملفات، جعلت التسجيلات النهائية داخل قواعد بيانات الويب تنقسم إلى قسمين ؛ **الأول :** مجموعة الحقول الببليوجرافية المكونة لشكل التسجيلة النهائية داخل قاعدة البيانات وهي حقول بيانات الوصف، **الثاني :** الملف الذي يحوي المعلومات النهائية، ويختلف في شكله من قاعدة إلى أخرى بحسب طبيعة المعلومات التي يحملها فهناك ملفات النصوص doc وملفات pdf التي تحوي البيانات النصية، أيضا هناك ملفات wmv, png, mp3... وغيرها من الأنواع التي تحوي المعلومات المصورة والمسموعة.

ولأن الملفات المصاحبة للتسجيلات الببليوجرافية لا يمكن عرضها أثناء عرض التسجيلات، فضلاً عن أنها تحتاج إلى برامج خاصة لتشغيلها، فقد تم وضعها كعنصر نصي داخل التسجيلة يحمل رابطًا link إلى ملف المعلومات النهائية. وقد عكست قواعد بيانات عينة الدراسة ذلك في وجود رابط إلى ملفات pdf في قاعدة eric، وملفات mp3 في قاعدتي allmusic, rhapsody. وملفات wmv في قاعدة movietome.

٤/٥/١/٤ البيانات المصاحبة للنتائج:

تمثل الوحدات المعروضة في واجهات عرض نتائج البحث الصورة الأولية والمصغرة للتسجيلات النهائية لقاعدة البيانات. وتحمل هذه الوحدات مؤشرًا على طبيعة ومحتوى التسجيلات النهائية. وتتكون وحدات النتائج من شقين أساسيين قد يجتمعان أو لا يجتمعان في عرض نتائج البحث، وهذان النوعان هما :

أولاً: بيانات الوصف : وهي البيانات التي تخص المحتوى الموضوعي لتسجيلات قواعد البيانات مثل بيانات عن المؤلف والعنوان ورؤوس موضوعات التسجيلة.

ثانيًا: خيارات العمل : وهي خيارات أو أوامر توفرها قاعدة البيانات للتعامل مع التسجيلة أو ملف المعلومات مثل خيارات التحميل والشراء والطباعة.

وقد كشفت عينة الدراسة عن أن قواعد البيانات ذات الطابع التجاري هي الأكثر حرصًا على عرض بيانات متنوعة من البيانات الوصفية وخيارات التعامل، وقد ظهر ذلك في قواعد بيانات amazon, corbis. كما أن هناك بعض قواعد البيانات التي لا تعطي أي بيانات في عرض النتائج، كقواعد البيانات النصية غير التجارية التي تعرض فقط رابطًا للنص أو الملف النهائي للمعلومات. أما خيارات العمل في قاعدة eric فلا تتجه إلى النواحي التجارية، وإنما تنحصر في خيارات الطباعة أو الإضافة للنتائج المفضلة للمستفيد.

٥/٥/١/٤ عدد الوحدات المعروضة:

تتيح بعض قواعد البيانات خيار عدد النتائج المعروضة في واجهات عرض النتائج، وتزداد أهمية هذا الإجراء إذا ما كانت مجموعة النتائج المعروضة تعتمد في اختيارها وانتقائها على المشاهدة. وقد انقسمت عينة الدراسة في ذلك إلى اتجاهات مختلفة فهناك قواعد بيانات لا تحدد عدد النتائج أمام المستفيد، ومنها ما يعرض عدد الوحدات المعروضة في كل صفحة، غير أنه لا يمكن للمستفيد التحكم في العدد المعروض، ومن قواعد البيانات ما يتيح إمكانية تحديد عدد الوحدات المعروضة لكل صفحة مثل قاعدة بيانات eric.

٦/٥/١/٤ البحث داخل النتائج:

تعتمد بعض قواعد بيانات الويب خدمة البحث داخل النتائج المسترجعة، وهو أسلوب يفيد في تضييق الجوانب الموضوعية للبحث؛ حيث يطلع المستفيد على مجموعات النتائج التي خرجت أمامه، مكونًا فكرة أفضل لإجراء البحث ثم البحث داخل النتائج المسترجعة للخروج بأكثرها إفادة لموضوعه. وتقدم هذه الخدمة غالبًا تحت مسمى تنقيح البحث أو refine search، واقتصر وجود هذه الخدمة على قاعدة البيانات eric التي قدمت هذه الخدمة تحت مسمى search within results. ويمكن القول إن هذه الخاصية تدرج دائمًا ضمن خدمات قواعد

بيانـات الويـب، غـير أن الباحـث آثـر ذكرهـا هنـا لارتباطهـا الوثيـق بطرائـق وخصائص عرض النتائج المسترجعة.

٧/٥/١/٤ عرض التسجيلات أو الروابط:

تهدف واجهات عرض نتائج البحث إلى إتاحة مختلف الوحدات الملائمة لموضوع البحث أمام المستفيد مجتمعة في ترتيب أو نسق عرض واحد، ومن ثم يستطيع المستفيد تصفح مختلف النتائج المسترجعة لاختيار وانتقاء أكثرها ملاءمة لموضوع بحثه. وتسلك قواعد بيانات الويب في عرض نتائج البحث سبلاً مختلفة؛ حيث يتجـه بعـضها إلى عـرض نتـائج البحـث بـشكل مفـصل في شـكل تـسجيلات مختصرة للتسجيلات الببليوجرافية النهائية، حتى يتمكن المستفيد من تحديد محتوى كل وحدة من الوحدات المسترجعة، والتعرف إلى ما بها من خصائص ومعلومات. ويتجه بعضها الآخر من قواعد بيانات الويب إلى عرض نتائج البحث في شكل مختصر يقتصر في أحيان كثيرة على عنوان ورابط أو مسار link or URL التسجيلات النهائية.

وقد كشف واقع عينة الدراسة من قواعد بيانات الويب على وجود رابط ما بين طبيعة وبناء قواعد البيانات وبين طرائق عرض نتائج البحث ؛ حيث تتجه معظم قواعد البيانات التجارية إلى عرض وصف مختصر في شكل تسجيلة ببليوجرافية مصغرة عن كل نتيجة داخل نتائج البحث، ومن أهم نماذجها amazon, corbis, rhapsody، أما قواعد البيانات الأخرى فتتجه إلى الاكتفاء بعرض عنوان ورابط التسجيلة النهائية أثناء عرض نتائج البحث، مثل قواعد medline, animal search. ويرى الباحث أن طريقة عرض نتائج البحث إنما هي انعكاس لقدرات قاعدة البيانات في بناء تسجيلاتها الببليوجرافية ؛ حيث يمكن لقاعدة البيانات إنتاج مستويات عدة مختصرة وكاملة للتسجيلة الببليوجرافية وعرضها أثناء عرض نتائج البحث، طالما اعتمدت من البداية على بناء تسجيلات نهائية ذات كفاءة ودقة عالية، وأهم ما يمثل ذلك في عينة الدراسة، قاعدة البيانات eric.

٦/١/٤ محتوى قواعد البيانات:

إن السبب الرئيس لاختفاء قواعد بيانات الويب عن محركات البحث وجعلها غير مرئية، هو المحتوى الداخلي لقواعد البيانات ؛ حيث تفتقر محركات البحث لآليات التعامل مع محتوى قواعد البيانات المتسم بالديناميكية، وسرعة التغيير، وتشتت الأشكال، وتعدد أنواع الملفات... وغيرها من الخصائص الأخرى التي تجعل من محتوى قواعد بيانات الويب كائنًا هلاميًا لا يمكن الوصول إليه أو التحكم به. ومن ثم اهتم الباحث بدراسة محتوى قواعد بيانات الويب، مركزًا على جوانب أهمها أشكال وأنواع ملفات قواعد البيانات، وكيفية الوصول إلى ملفات قواعد البيانات والبحث عنها، بالإضافة إلى كون ملف البيانات جزء من التسجيلة الداخلية أم لا. ويكفي القول بأن جانب محتوى مواقع الويب غير المرئية هو أساس التقسيم العام لفئات الويب غير المرئية، وهو ما قد تناولته مختلف الدراسات في إطار تشخيص طبيعة الويب غير المرئية ؛ حيث أوردت أنواعًا عدة للمحتوى غير المرئي منها ما يلي[1]:

- الويب ذات الملكية.
- الويب العميقة.
- الويب غير الشفافة.
- الويب غير المرئية الحقيقية.

وانطلاقًا من أن قواعد بيانات الويب الآن لا تحمل تسجيلات ببليوجرافية فقط، وإنما تحمل ملفات للمعلومات النهائية سواء أكانت نصوصًا كاملة أم وسائط متعددة، فإن التركيز الآن ينصب على طبيعة الملفات وطبيعة إدارتها داخل قواعد بيانات الويب، خاصة أنها ستكون الغاية الكبرى التي يسعى إليها محرك بحث قواعد البيانات بحثًا واسترجاعًا وعرضًا أمام المستفيدين. ويوضح الجدولان (٤– ١٠) و(٤– ١١) طبيعة بناء وإدارة ملفات المعلومات النهائية داخل عينة قواعد بيانات الويب كما يلي:

(١) Hobbins, Katie. Surfing the Invisible Web, 2001, cited at 10/10/2008, cited at www.sciencedirect.com

الجدول (٤- ١٠) المحتوى الداخلي لقواعد البيانات النصية

محتوى قواعد البيانات		قواعد نصية متخصصة				
		السيرة	التربية	الحيوان	الطب	التسوق
		biography	eric	Animal search	medlineplus	amazon
أنواع الملفات	شكل ثابت	---	√	√	√	--
	أشكال متعددة	√	--	--	---	√
البحث عن الملفات	داخل الملف	---	--	--	---	--
	النصوص المحيطة metadata	√	√	√	√	√
ملف المعلومات داخل التسجيلة	عرض كرابط	√	√	--	---	√
	عرض مباشر	---	--	√	√	--
نوع المحتوى	ثابت	√	√	√	√	--
	ديناميكي	---	---	---	---	√

الجدول (٤- ١١) المحتوى الداخلي لقواعد بيانات الوسائط المتعددة

محتوى قواعد البيانات		قواعد وسائط متعددة					
		الصوت		صور ثابتة		صور متحركة	
		allmusic	rhapsody	corbis	freefoto	movietome	world-of-movies
أنواع الملفات	شكل ثابت	---	---	---	---	√	√
	أشكال متعددة	√	√	√	√	---	---

البحث عن الملفات	داخل الملف	---	---	---	---	---	---
	النصوص المحيطة metadata	√	√	√	√	√	√
ملف المعلومات داخل التسجيلة	عرض كرابط	√	√	--	--	√	√
	عرض مباشر			√	√	---	---
نوع المحتوى	ثابت	---	---	√	√	---	---
	ديناميكي	√	√	--	--	√	√

تشير الجداول السابقة (٤- ٨) و(٤- ٩) إلى مجموعة من الملاحظات المهمة، التي يمكن شرحها فيما يلي :

١/٦/١/٤ أنواع الملفات:

أولاً: تذخر قواعد بيانات الويب بكثير من أشكال المعلومات وأنواع ملفات حمل المعلومات في وسائطها المختلفة، وذلك نتج عن إمكانية حمل وسيط المعلومات الواحد، سواء النصي أو الصوتي أو المرئي في عدة أشكال formats من الملفات. فعند إجراء بحث لموضوع ما يمكن استرجاع وسيط النص مصوغًا في عدة أشكال، مثل doc, pdf, html. كما أنه يمكن استرجاع المعلومات المصورة الثابتة في اكثر من شكل من الملفات مثل png, gif, jpeg ... إلخ.

ثانيًا: وقد عكست عينة الدراسة أيضًا أن اختلاف وسيط حمل المعلومات من قاعدة بيانات إلى أخرى هو سبب تنوع أو ثبات شكل الملفات التي تحويها قواعد بيانات الويب ؛ فقد تعددت أشكال الملفات داخل قاعدة البيانات المتخصصة في وسائط الصوت والصورة الثابتة، مثل قواعد corbis, rhapsody, allmusic, amazon، وهذا يرجع بطبيعة الحال إلى وجود أشكال متعددة من ملفات حمل الصوت والصورة، جاء تعددها نتيجة للإمكانات التقنية أو الفنية التي يتميز بها

كل شكل عن الآخر في جودة أو قدرة حمل المعلومات. وهذا المعيار يتجلى أكثر في ملفات حمل الصوت والصورة دون النصوص التي لا يهتم المستفيد كثيرًا بمدى حالة الجودة أو الوضوح لكتابة الكلمات النصية. وقد انعكس ذلك على قواعد بيانات النصوص، مثل medline, animal search & eric التي اكتفت باحتواء شكل واحد من ملفات النصوص هو شكل الأكروبات pdf في قاعدة Eric، وملفات الويب html في قاعدتي animal search & medline .

٢/٦/١/٤ البحث في الملف أو النصوص المحيطة metadata:

لقد أفرزت البيئة الرقمية الممثلة في الويب نوعًا أو أسلوبًا جديدًا من أساليب البحث عن المعلومات الرقمية، تمثل هذا الأسلوب فيما يسمى بأسلوب الاسترجاع المبني على المحتوى Content Based Retrieval CBR. وهو الأسلوب التقني المقابل لأسلوب البحث والاسترجاع المبني على الكلمات أو النصوص Text Based Retrieval TBR. وتكمن آلية عمل أسلوب البحث المبني على المحتوى في استخدام عناصر البنية texture والشكل shape واللون color لاسترجاع ملفات المعلومات المصورة، سواء تلك الثابتة أو المتحركة. وقد ظهر هذا المفهوم على يد Kato عام ١٩٩٢م عندما بدأ العمل في اتجاه استرجاع الصور، ويكون برنامج أو نظام الاسترجاع ذا خصائص تمنحه القدرة على تحليل مكونات الصورة الأساسية وهي اللون والشكل shape والبنية texture. ولا تدخل للنص هنا في الاسترجاع سواء بالوصف أو بالكلمات المفتاحية، وإنما يعتمد الأمر فقط على الخصائص والأساليب التقنية الخالصة. وهذا النوع من الاسترجاع تكون الحاجة له حيث يكون فيه استخدام الواصفات أو النص غير وارد مثل التعرف إلى بصمات شخص ما ومحاولة مطابقتها على بصمات أخرى في قاعدة بيانات. أو التعرف إلى شكل فيروس أو بكتيريا غير معروفة ومطلوب مطابقتها على مواد أخرى شبيهة في قاعدة بيانات(١).

(١) سيد ربيع سيد إبراهيم. محركات بحث الصور الثابتة : دراسة تحليلية.- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧م، ص ١٠٧.

لقد عكس واقع عينة الدراسة، خاصة قواعد الوسائط المتعددة، أن أسلوب البحث بالنص ما زال الأسلوب الغالب في استرجاع ملفات المعلومات الصوتية والمصورة على الويب، كما أن أدوات بحث الويب شاملة محركات البحث وقواعد البيانات المتخصصة في الصوت والصورة ما زالت تعتمد على استرجاع ملفاتها باستخدام النصوص أو عناصر الوصف الفني للملفات metadata. وهذا ما قد انعكس في افتقار كل قواعد بيانات عينة الدراسة النصية وغير النصية إلى اعتماد أسلوب الاسترجاع المبني على المحتوى، واستخدامها جميعًا لاسترجاع ملفاتها من خلال حقول البحث النصية، وبشكل آخر فإن الوصول للملفات قواعد البيانات لا يكون من خلال الملفات ذاتها، وإنما من خلال حقول التسجيلة الببليوجرافية التي ينتمي إليها ملف المعلومات.

٣/٦/١/٤ الملف كجزء من التسجيلة:

تتواجد ملفات المعلومات النهائية داخل قواعد البيانات التي تحوي أشكالاً مختلفة من المعلومات. أما قواعد البيانات النصية أو التي لا تحمل النصوص الكاملة فتحتوي فقط على تسجيلات ببليوجرافية مكونة من حقول الوصف الببليوجرافي. وانطلاقًا من أن الملفات بطبيعتها تحتوي على معلومات ذات وسيط أو شكل مختلف، فإنها دائمًا تنفصل عن حقول التسجيلة الببليوجرافية، مع وجود رابط لاستدعاء الملف وعرضه أمام المستفيد. وقد عكس واقع عينة الدراسة أن ملفات الصورة فقط هي التي يمكن عرضها أثناء عرض التسجيلة الببليوجرافية، متمثلة في أحد الحقول المكونة لهذه التسجيلة. وقد انعكس ذلك في قاعدتي corbis, freefoto. أما المعلومات المصورة المتحركة أو الصوتية فتعرض في ملف منفصل يختلف عن واجهات عرض التسجيلات الكاملة، وهو ما بدا واضحًا في عمل قواعد بيانات الموسيقى والأفلام. كما أن هناك بعض قواعد البيانات التي لا تعتمد على تسجيلات ببليوجرافية في معناها الكامل، وإنما تتجه لعرض ملفاتها النصية مباشرة، خاصة إذا كانت هذه النصوص في شكل صفحات ويب html files لا تحتاج إلى عرض منفصل أو برنامج مختلف.

٤/٦/١/٤ محتوى ثابت أم ديناميكي:

مكنت التقنيات الرقمية الحديثة قواعد بيانات الويب من القدرة على العمل بنمط تفاعلي ومتغير؛ حيث يمكن لقواعد بيانات الويب عرض محتوياتها الداخلية الآن في واجهة واحدة وان اختلفت المعلومات المعروضة في طبيعتها، كما يمكن لقاعدة بيانات الويب صياغة المحتوى الداخلي وتغييره بشكل ديناميكي يمكن المستفيد من تجميع معلومات مختلفة والاطلاع على موضوعات متعددة في آن واحد.

لقد شكل المحتوى الديناميكي أو التفاعلي لمواقع الويب عامة حجر عثرة أمام محركات بحث الويب التي تجد نفسها أمام محتوى سريع التغيير، لا يمكن نسخه أو التحكم به، وهذا العامل هو أم عوامل ظهور الويب غير المرئية. وقد لاحظ الباحث ثمة رابط بين طبيعة قواعد البيانات التجارية أو الوسائط المتعددة التي تسعى لجذب المستفيد ومحاولة التمييز بطرق العرض والإبهار أمام المستفيد، خاصة إن كانت المعلومات المعروضة يطغى عليها التنوع في جوانبها الموضوعية أو أشكالها المصورة والنصية. وقد تمثل ذلك في قواعد بيانات الصوت rhapsody, allmusic، وقواعد بيانات الأفلام movietome, world of movies، إلى جانب قاعدة البيانات التجارية amazon.

٧/١/٤ عرض التسجيلات النهائية:

تمثل التسجيلات الببليوجرافية نهاية الطريق الذي يسلكه المستفيد للوصول إلى المعلومات النهائية؛ حيث تنتهي عملية الإعداد الفني داخل قواعد بيانات الويب، بشقيها الوصف المادي والوصف الموضوعي، إلى تكوين شكل تسجيلة الوصف الببليوجرافي الكاملة. وتضم هذه التسجيلة ضمن حقولها الواصفات الموضوعية بجانب الوصف الموضوعي الشارح لموضوع ملف المعلومات. وتتعامل قواعد بيانات الويب مع التسجيلات الببليوجرافية بثلاثة مستويات للعرض أمام المستفيدين، وهي تختلف بين المستوى المختصر والكامل والفني ؛ حيث يتضمن

المستوى المختصر الحقول المهمة في التعريف بوسيط المعلومات، ويتضمن المستوى الكامل الحقول الكاملة، ويزيد على حقول الوصف بيانات أخرى مثل اسم المدخل أو تاريخ الإدخال ويشتمل عليها المستوى الفني في العرض.

ترجع أهمية الحديث عن التسجيلات النهائية هنا إلى أنها تحمل مختلف الكلمات الدالة أو حقول البحث التي يصل من خلالها المستفيد إلى وحدات المعلومات المطابقة لموضوع البحث. فضلاً عن أن هذه التسجيلات في مرحلة ما من عمل محرك بحث الويب غير المرئية، سيكون لها الدور الأكبر في التكشيف والتعريف بمحتوى قاعدة بيانات الويب. كما أنها ستكون المتحدث الأفصح أمام برامج الزاحف للتعبير عن مختلف التحديثات updates التي طرأت على محتوى قاعدة البيانات.

وسيتم تناول التسجيلات النهائية داخل قواعد بيانات الويب بالتركيز على جوانب مثل: مستويات عرض التسجيلة النهائية، وسرعة العرض أمام المستفيد، وطبيعة التسجيلة النهائية داخل قاعدة البيانات كتسجيلة ثابتة أو متعددة. ويوضح الجدولان (٤- ١٢) و(٤- ١٣) الخصائص المختلفة للتسجيلات النهائية داخل قواعد بيانات العينة.

الجدول (٤- ١٢) التسجيلات النهائية في قواعد البيانات النصية

(عرض التسجيلات النهائية)		قواعد نصية متخصصة				
		السيرة	التربية	الحيوان	الطب	التسوق
		biography	eric	Animal search	medlineplus	amazon
مستويات عرض التسجيلات	مستوى مختصر	---	--	--	---	---
	مستوى كامل	√	√	√	√	√
سرعة العرض وإتمام البحث		√	√	√	√	√
إنتاج التسجيلات الداخلية	تسجيلة مفردة	√	√	√	√	
	تسجيلات متعددة	---	--	--	---	√

الجدول (٤- ١٣) التسجيلات النهائية في قواعد بيانات الوسائط المتعددة

(عرض التسجيلات النهائية)		قواعد وسائط متعددة					
		الصوت		صور ثابتة		صور متحركة	
		allmusic	rhapsody	corbis	freefoto	movietome	world-of-movies
مستويات عرض التسجيلات	مستوى مختصر	---	---	---	---	---	---
	مستوى كامل	√	√	√	√	√	√
سرعة العرض وإتمام البحث		√	√	√	√	√	√
إنتاج التسجيلات الداخلية	تسجيلة مفردة			√	√	---	---
	تسجيلات متعددة	√	√	---	---	√	√

يتضح من الجدولين (٤- ١٢) و(٤- ١٣) بعض الملاحظات التي يمكن تلخيصها فيما يلي :

١/٧/١/٤ مستويات عرض التسجيلات:

تبنت مختلف قواعد بيانات عينة الدراسة عرض التسجيلة النهائية بشكل كامل يضم مختلف عناصر وحقول الوصف للمحتوى، وذلك في حالتي قواعد بيانات النصوص والوسائط المتعددة. إلى جانب ذلك فقد اعتمدت بعض قواعد النصوص على عرض المقالات أو النصوص الكاملة لها بدلا من التسجيلة النهائية، لأنها لا تدعم الشكل المحكم من التسجيلات الببليوجرافية وحقول

الوصف الفني، وإنما تحتوي فقط على المقالات النصية التي تخدم تخصص قاعدة البيانات، مثل ما جاء في قاعدتي animal search, medline. وكانت أفضل صور التسجيلات النهائية متمثلة في تسجيلات قاعدة البيانات eric التي عرضت نموذجًا للتسجيلة الببليوجرافية المكونة من حقول الوصف الفني، إلى جانب رابط للملف النصي في شكل pdf.

٢/٧/١/٤ سرعة العرض وإتمام البحث:

يعد جانب السرعة أحد المتطلبات الأساسية للمستفيدين في أداء مختلف مواقع الويب، ويقصد بالسرعة هنا القدرات التقنية لقواعد الويب في عرض بياناتها، وهو ما يتوقف بشكل مباشر على طبيعة المحتوى الداخلي لقاعدة البيانات وطبيعة وسيط حمل المعلومات داخلها ؛ حيث يختلف الوقت اللازم لتحميل صورة متحركة أو ثابتة عن الوقت المستنفذ في تحميل الملفات النصية. غير أنه لا يمكن إغفال الكثير من الجوانب الأخرى التي تملك تأثيرًا كبيرًا في سرعة التحميل أو العرض، ومن أهمها ؛ قدرات الحاسب الشخصي لدى المستفيد، وطبيعة تصميم واجهات قواعد البيانات. وقد عكس واقع عينة الدراسة قدرة مختلف قواعد بيانات عينة الدراسة على توفير المعلومات النهائية وإمكانات الوصول للتسجيلة الببليوجرافية في شكلها الكامل في أقصر وقت ممكن.

٣/٧/١/٤ تسجيلة مفردة أم تسجيلات متعددة:

تملك العديد من قواعد بيانات الويب الآن تقنيات لإنتاج واجهات مختلفة تعرض من خلالها التسجيلات الببليوجرافية، وساعد على ذلك انتشار الأشكال المختلفة من المعلومات المصورة المتحركة والثابتة التي تخطت بطبيعتها حاجز حقول الوصف النصي إلى إمكانات العرض المصور متعدد الجوانب. وقد عملت كثير من قواعد بيانات عينة الدراسة على تخصيص أكثر من واجهة أو صفحة ويب لعرض التسجيلة المختلفة، بحيث يعرض في كل واجهة واحدة جانب واحد فقط

من جوانب التسجيلة الببليوجرافية كلها، يحتوي كل جانب منها على مجموعة محددة متكاملة من حقول الوصف الببليوجرافي. وجاءت قاعدتا movietome, allmusic كأبرز نماذج عرض التسجيلة الواحدة في واجهات متعددة؛ فقد عرضت قاعدة allmusic بيانات عن المطرب في واجهة مستقلة، وأعماله الموسيقية في واجهة أخرى، وبيانات أخرى في واجهة ثالثة ... إلخ.

٢/٤ خاتمة:

تناول هذا الفصل بعض ملامح الواقع الذي تعمل فيه قواعد بيانات الويب؛ حيث تم تناول جوانب مختلفة في بحث واسترجاع معلومات تلك القواعد مثل؛ واجهة بحث قاعدة البيانات، وأساليب البحث المعتمد داخلها، ومستويات الوصول إلى المعلومات والخدمات المقدمة للمستفيدين. وجاء تناول التسجيلات الببليوجرافية النهائية داخل قواعد بيانات الويب لمحاولة رسم ملامح العمل أمام محرك بحث الويب غير المرئية، والإجراءات التي يمكن اتخاذها لتذليل مختلف الصعوبات التقنية أو الفنية التي يسير عليها واقع قواعد بيانات الويب غير المرئية. فمشكلة التغيير المتلاحق وقدرات إنتاج التسجيلات المتعددة وطبيعة الملفات ذات الأشكال المتباينة، قد شكلت عائقًا أمام محركات بحث الويب للتعامل مع محتوى قواعد البيانات، ومن ثم فإن التغلب عليها قد استلزم منا الاقتراب وملاحظة آليات العمل المختلفة. وهو ما عمل الباحث في هذا الفصل على تحقيقه من خلال دراسة واقع عينة قواعد بيانات الويب متعددة المجالات ومتباينة أشكال المحتوى الداخلي من المعلومات ووسائط حملها. بالإضافة إلى تركيز الدراسة التجريبية على خصائص مختلفة في قواعد بيانات الويب، يمكن أن تشكل هذه الخصائص فارقًا كبيرًا في طبيعة استرجاع محركات بحث الويب غير المرئية لمحتوى قواعد البيانات.

الفصل الخامس

محركات وأدلة الويب غير المرئية

الدراسة التجريبية

٥/٠ تمهيد :

تعتمد عمليات الوصول إلى محتوى قواعد بيانات الويب غير المرئية ومعالجته وتنظيمه على مدى التوافق في البنية بين قطبي الويب غير المرئية : قواعد البيانات ومحركات البحث. وقد تناول الفصل الرابع دراسة واقع عمل قواعد بيانات الويب وطبيعة بنيتها الداخلية والتسجيلات الببليوجرافية التي تكون مجموعات هذه القواعد. ومن ثم يمكن التعرف إلى أهم العوامل الداعمة لفتح محتوى هذه القواعد أمام محركات بحث الويب، بما يمكن معه الدخول لقواعد البيانات أمام برامج الزاحف. ويهدف الفصل الخامس إلى قراءة واقع عمل محركات بحث الويب غير المرئية. وقد رأى الباحث أثناء اختياره لعينة الدراسة من محركات وأدلة البحث السعي الحثيث من مصممي نظم محركات البحث لبناء محرك بحث قادر على تحليل وتكشيف الويب غير المرئية، إلا أن ذلك ما زال في طور البداية الذي لا يعطي الإمكانية الكاملة لمستفيدي الويب للوصول إلى محتوى الويب غير المرئية. ويحاول الباحث قراءة واقع محركات وأدلة بحث الويب غير المرئية، للوقوف على نقاط القوة والضعف في نماذج العمل الحالية، وذلك في سياق العناصر الثلاثة التالية :

الأول : يستقرئه الباحث من واقع عينة الدراسة وتفحصها بشكل عام، وهو يخص الحديث عن طبيعة محتوى الويب غير المرئية على الإنترنت، وخصائصه. وذلك كمقدمة لدراسة محركات وأدلة بحث الويب غير المرئية.

الثاني : يركز على محركات بحث قواعد البيانات، وواقع عمل برنامج الزاحف بها.

الثالث : فيتجه إلى بنية أدلة بحث قواعد بيانات الويب، وخصائص التقسيمات الموضوعية بها.

١/٥ خصائص تنظيم محتوى الويب غير المرئية :

لم يكن الباحث ليبدأ الحديث عن محركات وأدلة بحث الويب غير المرئية، حتى نستقرئ معاً واقع محتوى الويب غير المرئية ذاتها. التي تتكون من مزيج من الملفات الرقمية ذات الأنواع المتعددة، ومواقع ثابتة المحتوى أو ديناميكية، وقواعد بيانات في بنية مهيكلة محكمة high controled structure. وهو ما يمثل مقدمة الحديث عن آليات بحث هذه المواقع واسترجاع محتواها غير المرئي. إن نظم استرجاع الويب غير المرئية تتعدد في أشكالها وأنواعها كما تتعدد أشكال محتوى الويب غير المرئي ؛ فهناك محركات البحث التي تسعى وراء مواقع الويب التفاعلية، وهناك من المحركات يسعى وراء اختراق قواعد بيانات الويب غير المرئية، كما أن هناك من أدلة البحث ما يكتفي بتقديم قائمة من الروابط أو المسارات URLs التي تقدم للمستفيد فقط خدمة حصر مواقع الويب غير المرئية. ومن ثم فإن هذه الدراسة التجريبية لا تشمل جميع أشكال محتوى الويب غير المرئية، وإنما تقتصر فقط على محركات وأدلة بحث قواعد بيانات الويب التي تناولها الفصل الرابع.

يشتمل محتوى الويب غير المرئية على خصائص متنوعة يُنظر إليها من جوانب ثلاثة هي :

أولاً: إن محتوى الويب غير المرئية يصل إلى درجة من التعقيد بحيث لا يوجد محرك بحث يستطيع الوصول إلى هذا المحتوى واسترجاع ملفاته الداخلية، لأمور تتعلق بالملكية وقدرات برامج الزاحف، وطبيعة الملفات الرقمية داخل هذا المحتوى. وعلى ذلك فلم توجد حتى الآن أداة واحدة من أدوات البحث متمثلة في أحد المحركات أو الأدلة قادرة على التعامل مع هذا المحتوى وصولاً وتكشيفًا واسترجاعًا.

ثانيًا: يتشتت محتوى الويب غير المرئية بين ملفات رقمية مخالفة لطبيعة ملفات الويب html، ومواقع ذات محتوى ديناميكي، وقواعد بيانات ذات بنية مخالفة لطبيعة برامج الزاحف. ولذلك فقد سار الاهتمام باسترجاع المحتوى غير المرئي في اتجاهات غير معيارية ؛ مثل التركيز على تكشيف محتوى قواعد البيانات تارة، والتركيز على استرجاع ملفات doc, pdf, mdb...إلخ تارة أخرى دون تحديد تخصص تبنى على أساسه أدوات البحث.

ثالثًا : إن محتوى الويب لا يقتصر فقط على المحتوى النصي، وإنما يتعدى ذلك إلى الوسائط المتعددة، ومن ثم فإن أدوات بحث الويب غير المرئية يجب أن تفرد جزءاً منها للتركيز على ملفات الوسائط المتعددة داخل المحتوى غير المرئي.

وسوف يتم قراءة واقع هذه الجوانب الثلاثة من خلال واقع عينة الدراسة من محركات وأدلة البحث، التي تعكس طبيعة وخصائص المحتوى غير المرئي الذي تسعى الدراسة لمعالجته واسترجاعه. وسيتم النظر لأدوات البحث من خلال أنماطها في تنظيم محتوى الويب غير المرئي، وتركيزها على الملفات الرقمية أو قواعد البيانات، كذلك اهتمامها بالمحتوى النصي مقارنة بمحتوى الوسائط المتعددة. وذلك من خلال العناصر التالية:

١/١/٥ أنماط تنظيم مواقع الويب غير المرئية:

تمتلك الويب غير المرئية القدر الأكبر من حجم المعلومات الرقمية المتاحة على الويب كلها، غير أنها لا تملك القدر نفسه من أدوات البحث الذي تملكه الويب المرئية ؛ فمن اليسير لمستخدمي الويب استخدام أي من محركات أو أدلة البحث لاسترجاع محتوى مواقع وصفحات الويب، وهذا ليس هو الحال بالنسبة للويب غير المرئية التي ينقصها أدوات بحث لاسترجاع محتواها غير المرئي. ومن ثم فإن الويب غير المرئية في الوقت الحالي تعتمد على أنماط تنظيم عامة تجتهد فيها أدوات البحث لتنظيم واسترجاع المحتوى غير المرئي. وتتنوع هذه الأنماط بين الاعتماد

على البحث المباشر باستخدام الكلمات المفتاحية، وهنا تؤدي أدوات البحث دور وسيط البحث بين المستفيد وبين الويب غير المرئية، أو نمط الاعتماد على الترتيب والتقسيم الموضوعي لمواقع الويب غير المرئية، حتى يمكن الوصول إليها والبحث داخلها، وهنا تقوم أدوات البحث بدور الحصر والربط بمواقع الويب غير المرئية. وتوضح الجداول (٥-١ / أ، ب) أنماط التنظيم والاسترجاع كشفت عنها عينة الدراسة، وهي ممثلة في ثلاثة أساليب كما يلي :

الجدول (٥-١ / أ) أنماط تنظيم واسترجاع الويب غير المرئية في عينة الدراسة (المحركات)

أنماط تنظيم واسترجاع الويب غير المرئية	محركات بحث الويب غير المرئية					
	business	incywincy	infomine	scirus	scienceresearch	turbo10
التكشيف والبحث الحر	√	√	√	√	√	√
التقسيم الموضوعي والترتيب	---	√	√	---	√	---
قائمة الروابط المباشرة	√	√	---	--	√	---

الجدول (٥-١ / ب) أنماط تنظيم واسترجاع الويب غير المرئية في عينة الدراسة (الأدلة)

أنماط تنظيم واسترجاع الويب غير المرئية	أدلة بحث الويب غير المرئية					
	Academici-nfo	beaucoup	Complete-planet	Find-articles	Library-spot	Mag-portal
التكشيف والبحث الحر	---	√	√	√	√	√
التقسيم الموضوعي والترتيب	√	√	√	√	√	√
قائمة الروابط المباشرة	√	√	√	√	√	√

يشير الجدولان (٥-١ / أ ، ب) إلى أنماط تنظيم واسترجاع الويب غير المرئية في أدوات البحث وهي :

أولاً: تتمتع أدلة بحث الويب غير المرئية بالوفرة العددية وإمكانات البحث والاسترجاع للمحتوى غير المرئي ؛ حيث كان من الصعب على الباحث أثناء تجميعه لعينة الدراسة واستعراض أدوات البحث أن يجد محركات لبحث الويب غير المرئية بقدر ما وجد أدلة تحاول تنظيم واسترجاع المحتوى المرئي. وهذا ما تعكسه عينة أدوات البحث في الجدول رقم (٥-١ / أ)، حيث لا يشتمل على القدر نفسه من تواجد أساليب البحث، وهذا يوضح التركيز على بناء أدلة بحث الويب المرئية دون بناء المحركات. ويرجع الباحث ذلك إلى أنه من اليسير على مصممي نظم استرجاع الويب بناء قوائم وتقسيمات موضوعية لمواقع وقواعد بيانات الويب غير المرئية، بينما من الصعب عليهم بناء برامج الزاحف التي تستطيع الدخول إلى محتوى قواعد البيانات، وهذه البرامج تمثل بذاتها رأس حربة العمل لمحركات البحث.

ثانيًا: جاء محركا البحث incywinky & scienceresearch كأفضل محركات عينة الدراسة ؛ ويرجع ذلك إلى أن هذين المحركين قد صُمِما خصيصًا للعمل على تنظيم واسترجاع محتوى الويب غير المرئية بما له من آليات وخصائص يختلف بها عن محتوى مواقع الويب المرئية، ومن ثم تزويدهما بأساليب البحث الملائمة للتعامل مع الويب غير المرئية، خلاف محركات البحث الأخرى. وقد عكس ذلك أمران : الأول : المسمى الصريح لهذه المحركات بأنها محركات بحث الويب غير المرئية، حيث حمل أحدهما اسم the invisible web search engine. والثاني : أن أحد محركي البحث قد صُمم بواسطة شركة deep web technology ، التي تهتم بتقنيات بحث الويب غير المرئية.

ثالثًا: تنتهج مختلف أدوات بحث الويب الحالية منهجًا واحدًا في استرجاع المحتوى غير المرئي، وهذا المنهج مؤداه أن تحصر أدوات البحث مختلف مواقع الويب وقواعد البيانات ثم تعمل على وضع عناوينها في قوائم مرتبة هجائيًا

كروابط للمواقع ذاتها، وهذا أضعف الحلول، أما أقواها فأن تقوم أدوات البحث بالتعرف إلى محتوى مواقع الويب غير المرئية وتقسيمها إلى مجالات موضوعية عامة ثم موضوعات أكثر تخصصًا، وصولاً إلى المواقع ذاتها. ومن ثم كانت أدلة بحث الويب غير المرئية هي الأكثر انتشارًا واستخدامًا من محركات البحث حتى الآن.

رابعًا: جاء محركا البحث turbo10 & scirus كأضعف أدوات بحث عينة الدراسة احتواءً على أساليب لبحث الويب غير المرئية، وهذا يرجع إلى أن الطبيعة الحقيقية لهذين المحركين تتمثل في قربهما إلى محركات البحث المتعددة metasearch engines، وليس إلى محركات بحث الويب غير المرئية. وهذا يأخذنا إلى أن أدوات بحث الويب غير المرئية ما زالت غير ثابتة على مدلول محدد ومفهوم أدائي قاطع، يحول دون تداخلها مع الأنواع الأخرى من أدوات بحث الويب المرئية وغير المرئية. وهو ما سيتم الحديث عنه لاحقًا.

٢/١/٥ الويب غير المرئية النصية:

تختص الويب عامة دون غيرها من أجزاء الإنترنت الأخرى (جوفر gopher، آركي archie، نقل الملفات FTP ... إلخ) باحتواء مختلف وسائط حمل المعلومات الرقمية من نصوص ووسائط متعددة. ومن ثم فإن محاولات تنظيم واسترجاع الويب غير المرئية لابد أن تشمل شقي المحتوى النصي وغير النصي. ويكشف هذا الجزء مدى الاهتمام بالمحتوى غير النصي تنظيمًا واسترجاعًا داخل عينة الدراسة من أدوات البحث، وذلك من خلال تتبع أساليب البحث وآليات العمل التي تقوم عليها عينة الدراسة، للنظر إلى ما توفره هذه الأدوات من آليات لاسترجاع المحتوى غير النصي. ويوضح الجدولان (٥-٢ / أ، ب) واقع اهتمام أدوات بحث الويب غير المرئية بملفات الوسائط المتعددة من خلال تتبع خصائص الاهتمام المتمثلة في أساليب البحث التي توفرها هذه الأدوات لاسترجاع المحتوى غير النصي، بالإضافة إلى طبيعة المعلومات المسترجعة، سواء أكانت نصًا كاملاً أم وسائط

متعددة، وكان النص جزءا منها. ويصدق هذا كله أو ينفيه المجال والتخصص الشكلي لأدوات بحث الويب غير المرئية.

الجدول (٥-٢ / أ) استرجاع المحتوى غير النصي في عينة الدراسة (المحركات)

استرجاع المحتوى غير النصي	محركات بحث الويب غير المرئية					
	business	incywincy	infomine	scirus	scienceresearch	turbo10
طبيعة أساليب البحث	نصية	نصية	نصية	نصية	نصية	نصية
المعلومات المسترجعة	نصية	مزيج	نصية	نصية	نصية	نصية
المواقع موضع الاهتمام	نصية	مزيج	نصية	نصية	نصية	نصية
مجال التخصص	متخصص	عام	متخصص	متخصص	متخصص	عام

الجدول (٥-٢ / ب) استرجاع المحتوى غير النصي في عينة الدراسة (الأدلة)

استرجاع المحتوى غير النصي	أدلة بحث الويب غير المرئية					
	Academici-nfo	beaucoup	Complete-planet	Find-articles	Library-spot	Mag-portal
طبيعة أساليب البحث	نصية	نصية	نصية	نصية	نصية	نصية
المعلومات المسترجعة	نصية	مزيج	مزيج	نصية	مزيج	نصية
مواقع موضع الاهتمام	نصية	مزيج	مزيج	نصية	نصية	نصية
مجال التخصص	متخصص	عام	عام	عام	عام	عام

يوضح الجدولان (٥-٢ / أ، ب) بعض من حقائق تنظيم الويب غير المرئية في أدوات البحث هي :

أولاً: إن بداية العمل على تنظيم واسترجاع محتوى الويب غير المرئية، تشبه إلى حد كبير بداية التعامل مع الويب المرئية من مواقع وصفحات. ويمتد هذا التشابه إلى بداية كل من أدوات بحث الويب المرئية مع غير المرئية؛ حيث تعكس أدوات بحث عينة الدراسة أن جملة الاهتمام تنصب الآن على تنظيم واسترجاع المحتوى

النصي غير المرئي، مع الإهمال الجزئي للوسائط المتعددة داخل الويب غير المرئية. فقد بدأت محركات بحث الويب بالعمل على تنظيم وبحث الصفحات والمواقع التي تحوي المعلومات النصية، ثم خرج بعد ذلك جيل جديد من محركات البحث يهتم بتنظيم ومعالجة المواد المصورة والمسموعة.

ثانيًا: إن مجمل أساليب البحث التي تعتمد عليها عينة الدراسة انحصرت فقط في تلك التي تركز على التعامل مع النصوص والكلمات، ولم يرد أي من الأساليب المختصة بالتعامل مع المحتوى الوسائط المتعددة. فقد وردت أساليب البحث الحر بالكلمات المفتاحية، وأسلوب التصفح بالمصطلحات ورؤوس الموضوعات، والحصر والترتيب لعناوين مواقع الويب غير المرئية. وقد خلت عينة الدراسة تمامًا من الإشارة إلى إمكانية التعامل مع المواد غير النصية، سوى ما جاء في محرك البحث incywinky، ودليل البحث Complete-planet من إمكانية التعامل مع قواعد بيانات الأفلام والصور، وهو ما يمثل قدرًا ضئيلاً من الاهتمام بالوسائط المتعددة بين جملة أدوات بحث عينة الدراسة.

ثالثًا: إن ثمة رابط بين المجال الموضوعي المتخصص وبين الاهتمام بالمواد النصية دون الوسائط المتعددة ؛ فقد جاءت أغلب محركات البحث متخصصة في مجالها الموضوعي، كما أنها جاءت أيضًا مركزةً على المواد النصية في الغالب منها. وعلى الجانب الآخر فقد جاءت أغلب أدلة البحث عامة في مجالها الموضوعي، وارتبط ذلك بتبني العدد الأكبر منها للبحث في وسيطي النص والصورة معًا. مما يعني أن محركات بحث الويب غير المرئية ما زالت تعمل بشكل منحصر أكثر مما يجب أن تكون عليه، وفي المقابل فإن أدلة البحث أكثر عمومية مما يجب أن تكون عليه. ومن ثم يمكن القول إن هناك ارتباطًا جمع بين كلمة عام في مربع التخصص وكلمة مزيج في مربع المعلومات موضع الاهتمام، كما أن هناك ارتباطًا بين كلمة متخصص في المجال الموضوعي وكلمة نصي في المعلومات موضع الاهتمام.

رابعًا: يرى الباحث أنه على الرغم من اهتمام أدلة البحث بذكر روابط الوسائط المتعددة إلا أنها لم توفر لها آليات البحث الملائمة للتعامل مع هذه الوسائط، واكتفت أدلة بحث الويب المرئية في تعاملها مع الوسائط المتعددة بأن تضع قوائم وروابط تحوي عناوين أو مسارات الحصول على ملفات الصورة والصوت. ولا يرى الباحث في ذلك نموذجا لآلية مُحكَمة لاسترجاع معلومات الصوت والصورة داخل الويب غير المرئية، سوى أنها دليل حصري أو ثبت بهذه المعلومات.

٣/١/٥ الملفات النهائية وقواعد البيانات:

إن الويب غير المرئية لا تحمل فقط معلومات متعددة الأشكال وحسب، وإنما تحمل أيضًا أنواعًا مختلفة من المعلومات تبدأ من الملفات الرقمية ذات الصيغة الخاصة مثل PDF, DOC، مرورًا بمواقع الويب ذات المحتوى الديناميكي، وصولاً إلى قواعد البيانات التي تحوي بداخلها تسجيلات وملفات ذات بنية مهيكلة في جداول وملفات وحقول. وعلى ذلك فإن بناء أدوات لبحث الويب غير المرئية يجب أن يأخذ في الاعتبار مستوى المعلومات التي ستصل إليها أداة البحث وتتعامل معها تنظيما واسترجاعا، فيمكن أن تقف أداة البحث فقط عند مستوى استرجاع الملفات الرقمية ؛ مثل محاولات محرك البحث google في الوصول إلى ملفات pdf واسترجاعها. أيضا يمكن أن تتعدى أداة البحث كافة المستويات لتصل إلى المحتوى الداخلي لقواعد بيانات الويب غير المرئية، متخطية بذلك حدود واجهات البحث، والجداول المغلقة حتى تتمكن من استرجاع تسجيلاتها وملفاتها الداخلية. ويوضح الجدولان (٥-٣ / أ، ب) طبيعة مستويات الوصول لمحتوى الويب غير المرئية داخل عينة الدراسة من أدوات البحث، التي تنقسم في ذلك إلى أدوات تقتصر على استرجاع الملفات وأدوات أخرى تصل إلى عمق قواعد البيانات.

الجدول (٥-٣ / أ) مستويات الوصول لمحتوى الويب غير المرئية في عينة الدراسة (المحركات)

مستويات الوصول داخل أدوات البحث	محركات بحث الويب غير المرئية					
	business	incywincy	infomine	scirus	scienceresearch	turbo10
استرجاع الملفات	- - -	√	- - -	√	√	- - -
استرجاع المواقع	√	√	√	√	√	√
استرجاع محتوى قواعد البيانات	√	√	√	√	√	√

الجدول (٥-٣ / ب) مستويات الوصول لمحتوى الويب غير المرئية في عينة الدراسة (الأدلة)

مستويات الوصول داخل أدوات البحث	أدلة بحث الويب غير المرئية					
	Academici-nfo	beaucoup	Complete-planet	Find-articles	Library-spot	Mag-portal
استرجاع الملفات	---	---	---	√	---	√
استرجاع المواقع	√	√	√	√	√	√
استرجاع محتوى قواعد البيانات	---	---	√	√	---	√

يوضح الجدولان (٥-٣ / أ ، ب) مستويات الوصول إلى محتوى الويب غير المرئية بالنسبة لأدوات البحث، وهي :

أولاً: تعمل أدوات بحث الويب غير المرئية في استرجاع المحتوى غير المرئي على تبني مستويات ثلاث هي ؛ **الأول** : مستوى استرجاع الملفات، وهي أعمق نقطة يمكن أن تصل إليها أدوات بحث الويب غير المرئية ؛ حيث تحوي الملفات النهائية غاية المعلومات التي يحتاجها المستفيد، التي تتمثل في النصوص الكاملة أو الصور الرقمية المتحركة أو الثابتة، غير أن هذه الملفات لا تنتمي إلى قواعد بيانات الويب، وإنما تنتمي إلى مواقع تفاعلية أو ذات محتوى متغير.

الثاني: مستوى استرجاع مواقع الويب، ويعني اكتفاء أدوات البحث بعرض عناوين مواقع الويب غير المرئية، حتى يتمكن المستفيد من الوصول إليها والبحث بها مباشرة، وهو دور الوسيط فقط أو الرابط بين المستفيد ومواقع الويب غير المرئية.
الثالث : مستوى استرجاع محتوى قواعد البيانات، وفي هذا المستوى تعمل أدوات البحث على اختراق أبواب قواعد البيانات واسترجاع محتواها من التسجيلات الداخلية أو تخطي ذلك إلى المستوى الأول وهو استرجاع الملفات النهائية داخل قاعدة البيانات. وإذا ما نجحت محركات وأدلة البحث في الوصول إلى هذه الملفات تمكنت بالفعل من كشف المحتوى غير المرئي وعرضه أمام مستخدمي الويب.

ثانيًا: حاولت محركات بحث عينة الدراسة الوصول مباشرة إلى ملفات المعلومات النهائية، سواء أكان ذلك على مستوى ملفات مواقع الويب أم على مستوى ملفات قواعد بيانات الويب. وهي بذلك تتفوق في بحث الويب غير المرئية على أدلة البحث التي تكتفي فقط بالحصر والترتيب لمواقع الويب غير المرئية ساعية إلى الربط بين المستفيد وبين الويب غير المرئية بواسطة مساراتها UTLs.

ثالثًا: تكشف النتائج الخاصة بمستوى استرجاع محتوى قواعد البيانات مدى الفارق بين محركات وأدلة البحث في السعي إلى كشف ظلمة الويب غير المرئية، ومن ثم فإننا إذا ما احتجنا لبناء أفضل أدوات بحث الويب غير المرئية تنظيمًا واسترجاعًا، فيجب التركيز على بنية محركات البحث وآليات عملها. وهو ما تهدف إليه هذه الدراسة. أما ما تقوم به أدلة البحث من ترتيب وحصر لمواقع الويب، فهو المفهوم القاصر الذي تتبناه أدوات البحث الحالية، أو يمكن القول بأنه أقصى ما تملكه أدوات بحث الويب غير المرئية حتى الآن في تعاملها مع استرجاع المحتوى غير المرئي لقواعد البيانات.

٢/٥ محركات بحث قواعد البيانات:

ارتبطت محركات البحث منذ نشأة الويب بالعمل على تنظيم واسترجاع صفحات ومواقع الويب، وذلك من خلال استخدام ثلاثة برامج متكاملة هي

الزاحف والمكشف وآليات البحث، يختص فيها الزاحف بالسعي خلف الروابط الفائقة ومسارات الويب حتى يجمع أكبر قدر ممكن من صفحات ومواقع الويب ثم نسخها إلى المكشف، الذي يعمل بدوره على تحليل محتوى هذه الصفحات والمواقع والتعرف إلى الكلمات الكشفية بها حتى يمكن استرجاع هذه المواقع ثانيةً تبعًا للكلمات الكشفية، وذلك باستخدام آليات البحث التي تتشكل من خلالها استراتيجيات البحث داخل قاعدة محرك بحث الويب. وهذا الوصف المختصر لعمل محرك بحث الويب المرئية لا يختلف إطلاقًا عن وصف العمل داخل محرك بحث الويب غير المرئية، سوى التقنيات والآليات التي تتلاءم مع قواعد بيانات الويب، وذلك لكون هذه القواعد تختلف في بنيتها وتنظيمها عن مواقع وصفحات الويب html files.

وتتشابه آلية عمل محركات بحث الويب غير المرئية مع ما تقوم به محركات البحث المتعددة، غير أن محركات بحث قواعد البيانات تحتاج إلى تقنيات أعلى وأكثر إحكامًا؛ لتعاملها مع قواعد البيانات. والدليل على ذلك أن أدوات بحث الويب غير المرئية تحتاج إلى التوقف عند تحديد ماهيتها ووظيفتها التي ستقوم بها على الويب، وهناك خلط في المفهوم بين محركات بحث الويب غير المرئية وبين محركات البحث المتعددة، ومثال لذلك وجود محرك البحث turbo10 بين أدوات بحث الويب غير المرئية، في حين أنه عمليًا محرك بحث متعدد يبحث في محركات البحث الأخرى.

ويسعى الباحث في هذا الجزء التطبيقي لمحركات بحث الويب غير المرئية إلى دراسة آليات العمل داخل الجيل الجديد المصاحب لميلاد الويب غير المرئية، مستخدمًا في ذلك وصف البرامج التي تعمل بها محركات الويب. وسوف ينصب التركيز على الجوانب المؤثرة في عمل محرك بحث الويب غير المرئية المقترح لهذه الدراسة، ومن هذه الجوانب التعرف إلى خصائص واجهات البحث التي ستكون الواجهة العامة لبناء استراتيجية البحث المزمع إرسالها بعد ذلك إلى مختلف قواعد

بيانات الويب لبحث محتواها ، أيضًا التركيز على برنامج الزاحف ومحاولة رصد الإشارات الدالة على نجاحه أو إخفاقه في التعامل مع محتوى قواعد بيانات الويب والمشكلات التي تتعثر فيها برامج الزاحف. بعد ذلك يأتي تناول واجهات عرض النتائج والخدمات التي تقدمها محركات بحث الويب غير المرئية لاستكمال شكل نظام استرجاع الويب غير المرئية. وتتمثل خصائص دراسة واقع محركات بحث الويب غير المرئية في العناصر التالية:

- واجهات بحث محركات بحث الويب غير المرئية:

١. نموذج البحث.

٢. أساليب البحث.

٣. آليات البحث المستخدمة.

٤. حقول البحث الببليوجرافية.

- برامج زحف محركات بحث الويب غير المرئية:

٥. مسح روابط الويب.

٦. التعامل مع بحث قواعد البيانات.

٧. تكشيف قواعد البيانات.

- واجهات عرض النتائج المستدعاة من قواعد البيانات:

٨. ترتيب نتائج البحث.

٩. خصائص الفرز والاستبعاد.

١٠. أنماط عرض الوحدات المسترجعة.

- خدمات محرك بحث الويب غير المرئية.

وسيتم تناول هذه العناصر كالتالي:

١/٢/٥ واجهات بحث محركات بحث الويب غير المرئية:

تتمتع واجهة البحث داخل أدوات البحث بقدر كبير من الأهمية قد تفوق به بعض المكونات الرئيسية لأداة بحث الويب. وتتمثل هذه الأهمية في أنها أول ما يجد المستفيد من تفاعل بينه وبين أداة البحث، فضلاً عن أنها منفذ وسبيل للتعبير عن الحاجة الموضوعية والمعلومات التي يبتغيها المستفيد. وإذا ما نجح المستفيد في إدراك الجوانب الوظيفية لواجهة البحث، كان بحثه عن المعلومات مثمرًا بما يتصل بموضوع البحث. وإذا كانت درجة الأهمية لواجهة البحث في أدلة ومحركات بحث الويب المرئية كبيرة، فإن أهمية واجهة البحث ستكون أكبر عند الحديث عن واجهات محركات بحث قواعد بيانات الويب؛ حيث ستكون واجهة البحث هنا بديلاً عن واجهات عدة تقدر بعدد قواعد بيانات الويب التي يصل إليها محرك البحث، حيث يجب أن تبلغ من التوافق مع واجهات قواعد بيانات الويب الدرجة التي تجعل المستفيد قادرًا على إرسال كلمات البحث المفتاحية إلى مختلف قواعد البيانات في آن واحد.

إن الدور الذي تقوم به واجهة محرك بحث قواعد بيانات الويب غير المرئية هو ذاته الدور الذي تقوم به واجهة محرك البحث المتعدد؛ حيث تستقبل واجهة محرك البحث المتعدد كلمات الاستفسار ثم توجهها إلى محركات البحث ذات الصلة، ثم تتهيأ مستقبلة مجموعات النتائج التي تم استرجاعها عن طريق كل محركات البحث. وبالمثل فإن محرك بحث الويب غير المرئية سيتعامل مع أكثر من قاعدة بيانات في آن واحد، يرسل إليها استفسارات البحث ثم يستقبل منها مجموعة النتائج، ومن ثم يعيد فرزها وترتيبها لعرضها أمام المستفيد. غير أن واجهة محرك بحث الويب غير المرئية المقترح في هذه الدراسة سوف تتعامل مع

واجهات بحث قواعد بيانات الويب التي تملك آليات بحث تفوق الآليات التي تعتمد عليها محركات بحث الويب، فضلاً عن أن قواعد بيانات الويب تعتمد على استخدام بنية التسجيلات الببليوجرافية التي يمكن أن ينتج عنها حقول نشطة للبحث توضع في واجهة قاعدة البيانات حتى يتم استخدامها في استرجاع التسجيلات النهائية. ويسعى الباحث في هذا الجزء إلى دراسة واقع واجهات بحث محركات الويب غير المرئية الحالية من خلال مجموعة العناصر لها، مثل : نموذج البحث، وآليات البحث التي تدعمها محركات بحث قواعد البيانات، إلى جانب مجموعة الحقول البحثية التي توجد في واجهات البحث ومدى علاقتها بما يمكن أن تكون عليه حقول بحث قواعد البيانات. وذلك لأجل تحقيق أكبر قدر من التوافق.

١/١/٢/٥ شكل واجهة البحث وبناؤها:

تتبنى أدوات بحث الويب أحد أسلوبين عند بنائها لواجهات البحث ؛ فإما أن تعتمد على واجهة بسيطة تحمل فقط فراغ البحث وبعض الإرشادات والجوانب التفصيلية البسيطة عن أداة البحث، وإما أن تعتمد على واجهات بحث مركبة تتكون من تفاصيل بحث متعددة توضع أمام المستفيد لتوضح مختلف خيارات وحقول البحث. غير أن هذا الاختيار بين الواجهتين البسيطة المباشرة والمركبة تعتمد أولاً وأخيرًا على طبيعة آليات البحث ونطاق التخصص الشكلي والموضوعي لمحرك البحث، وهنا فإن واجهة محركات بحث قواعد البيانات تحتاج إلى التحلي بعناصر وخيارات بحث متعددة، فضلاً عن جوانب أخرى تكميلية لعملية البحث مثل تحديد القاعدة أو قواعد البيانات موضع الاهتمام. فضلاً عن أن محركات بحث قواعد البيانات تتعامل مع وحدات ذات بنية معقدة وتركيب محكم هي قواعد البيانات. ويوضح الجدول (٥-٤) شكل واجهة محركات بحث قواعد البيانات عينة الدراسة.

الجدول (٥-٤) شكل واجهة محرك بحث قواعد البيانات

أساليب البحث	محركات بحث الويب غير المرئية					
	business	incywincy	infomine	scirus	scienceresearch	turbo10
واجهة بسيطة مباشرة	√	---	---	---	√	√
واجهة معقدة مركبة	---	√	√	√	---	---

يشير الجدول (٥-٤) إلى مجموعة من الملاحظات هي:

أولاً: إن ثمة ارتباطًا في واجهات محركات البحث، ألا وهو أن الواجهات البسيطة المباشرة توجد في أدوات بحث الويب غير المرئية التي تميل إلى كونها أدلة بحث أكثر من كونها محركات بحث ؛ حيث تحتوي واجهات البحث المباشرة على فراغ لكتابة الكلمات البحثية ثم ثبت أو قائمة بالمصطلحات أو عناوين المواقع التي تبحث داخلها، مما يعني أنها لا تملك آليات البحث المتقدمة التي توفر البحث بالكلمات المفتاحية، وبذلك فهي تسير في طريق أدلة بحث الويب غير المرئية.

ثانيًا: جاءت واجهة محرك incywinky كأفضل واجهة بحث تحتوي على عناصر بحث متكاملة، وعلى الجانب الآخر فقد كانت واجهة محرك infomine هي أفضل الواجهات تركيبًا وملاءمة لبحث قواعد البيانات باحتوائها على حقول البحث الببليوجرافية، ثم كانت واجهة scirus هي الأبسط في بنيتها بين مختلف المحركات التي تحتوي على واجهات مركبة.

٢/١/٢/٥ أساليب البحث:

تتنوع آليات وأساليب البحث على الويب عامة بين أسلوبين أساسيين هما البحث الحر بالكلمات المفتاحية والبحث بالتصفح. ويندرج تحت كل منهما

مختلف الآليات التي تعتمد عليها أدوات بحث الويب المرئية وغير المرئية ؛ فعلى سبيل المثال، سواء اعتمدت أداة البحث على التقسيم الموضوعي بالمصطلحات أو الأشكال المصورة مثالاً أو التصفح من خلال الوحدات والتسجيلات ذاتها ، فإن كل ذلك يسمى في النهاية بأسلوب البحث بتصفح المعلومات. كما أن استخدام الحقول الببليوجرافية أو آليات البحث البولياني مباشرة أو اعتماد آليات أخرى تتكون من كلمات مفتاحية ، فإن كل ذلك يسمى نهاية بأسلوب البحث الحر بالكلمات الدالة.

إن الاختلاف بين أسلوبي البحث الحر والبحث بالتصفح قد أثر في عمل أدوات بحث الويب والمستفيدين منها من نواحٍ عدة منها :

أولاً : إن أسلوب البحث بالتصفح يستخدم في الغالب من جانب المستفيدين ذوي مهارات البحث المحدودة ، كما أن أدوات بحث الوسائط المتعددة تعتمد على أسلوب التصفح في حالات عدة ، خاصة عند استرجاع نتائجها أمام المستفيدين.

ثانيًا : تجد أدوات بحث الويب غير المرئية في أسلوب البحث بالتصفح الحل الأمثل والأسرع للتعامل مع المحتوى غير المرئي ، حيث إنها ما زالت لا تجيد أسلوب البحث الحر بالكلمات المفتاحية وآلياته التي تمكنها من استرجاع محتوى الويب غير المرئية.

ثالثًا : تعمل أدوات بحث الويب غير المرئية الحالية على تجميع عناوين قواعد بيانات الويب والمواقع التفاعلية ، ثم ترتيبها بحسب قطاعات موضوعية عامة أو متخصصة ، ومن ثم يمكن للمستفيد تصفحها واختيار مواقع أو قواعد البيانات الملائمة لموضوع البحث ، وعلى ذلك فإن أسلوب التصفح هو الأيسر بناءً واستخدامًا داخل أدوات بحث الويب غير المرئية ، في مقابل أسلوب البحث الحر الذي يتطلب الأمر معه تقنيات ذات كفاءة عالية وآليات بحث متوافقة مع تلك المستخدمة في مواقع وقواعد بيانات الويب غير المرئية.

الجدول (٥-٥) أساليب البحث داخل محركات بحث الويب غير المرئية عينة الدراسة

أساليب البحث		محركات بحث الويب غير المرئية					
		business	incywincy	infomine	scirus	scienceresearch	turbo10
البحث الحر		√	√	√	√	√	√
البحث بالتصفح	بالمصطلحات	---	√	√	---	√	---
	عنوان الموقع	√	√	---	---	---	---

يتضح من الجدول (٥-٥) مجموعة من الملاحظات هي:

أولاً: اتفقت جميع محركات عينة الدراسة في الاعتماد على أسلوب البحث الحر، بينما اختلفت في استخدام أسلوب البحث بالتصفح ؛ وذلك يرجع إلى أن الأسلوب الأسرع والأسهل في البحث يكمن في إدخال المستفيد الكلمات الدالة على موضوع بحثه وكتابتها في فراغ البحث. وعلى الرغم من اتفاق جميع المحركات على اتباع أسلوب البحث الحر، إلا أنها لم تتفق في آليات البحث المصاحبة لأسلوب البحث الحر ؛ حيث لم ترد أي من هذه الآليات على الإطلاق في بعض محركات العينة.

ثانيًا: اتفقت محركات البحث التي اعتمدت على التصفح بالتقسيم الموضوعي في تقسيم مجموعاتها إلى قطاعات عامة مثل المواد الترفيهية والعلمية والصحة ... إلخ، أو قطاعات عامة مثل العلوم الاجتماعية والإنسانيات والعلوم التطبيقية ... إلخ. ويدل ذلك على عدم اتباع أي من محركات بحث العينة لنظام مقنن مثل خطط التصنيف أو قوائم رؤوس الموضوعات. وقد احتوى محرك البحث infomine فقط على قائمة مرتبة هجائية بالمصطلحات المتخصصة في المجالات العلمية.

ثالثًا: جاء محرك البحث incywinky كأفضل محركات عينة الدراسة في جمعه بين العمل بأسلوب البحث الحر وتوفير آليات البحث الملائمة لذلك، وبين العمل بأسلوب التصفح بنوعيه ؛ التصفح بالمصطلحات الموضوعية، والتصفح

بعناوين مواقع الويب غير المرئية. بينما كان محرك البحث turbo10 أقل محركات بحث العينة في إمكانات العمل بأساليب البحث، حيث يعمل فقط بأسلوب البحث الحر، فضلاً عن غياب آليات البحث الداعمة لهذا الأسلوب.

٣/١/٢/٥ نموذج البحث داخل واجهة المحرك:

يبنى نموذج البحث في أدوات بحث الويب عامة داخل واجهات تلك الأدوات. ويتكون نموذج البحث من مجموعة من عناصر وحقول بحث تتكامل معًا لاسترجاع المعلومات؛ حيث تستقبل هذه العناصر الكلمات المفتاحية بحسب توزيعها بين حقول المؤلف والموضوع والشكل ... إلخ، ثم توجه معًا كاستراتيجية بحث تهدف إلى استرجاع معلومات ذات سمات محددة تتفق واستفسار البحث. ويتكون نموذج البحث من مكونين رئيسين هما : أولاً: مسمى عنصر البحث، وهو يدل على طبيعة المكان الذي ستصل إليه الكلمة الدالة مثل مسمى حقل المؤلف، حقل العنوان، حقل الموضوع، أما المكون الثاني فهو عنصر البحث ذاته، وقد يكون عبارة عن فراغ نصي text box لتلقي كلمات البحث، أو عنصر تحديد خصائص البحث chek box . وتتدرج نماذج بحث أدوات الاسترجاع على الويب بداية من النماذج البسيطة التي تقتصر على فراغ البحث لاستقبال الكلمات المفتاحية، ثم الانتهاء بنماذج البحث المتقدمة المعقدة التي تتكون من عناصر بحث مختلفة الوظيفة والأداء مثل؛ فراغات الكلمات المفتاحية، وأزرار الخيارات chek box وأزرار التحديد radio buttons. ويمكن القول إن أدوات بحث الويب المرئية ما زالت تعتمد في بنائها على نماذج البحث البسيطة التي يسهل على المستفيد استخدامها، التي تقتصر في أغلب الأحيان على فراغ كتابة الكلمات البحثية، وذلك ظنًا بأن أغلبية مستخدمي الويب لا يملكون مهارات البحث الكافية للتعامل مع نماذج البحث المعقدة المركبة.

إن عمل نموذج بحث محرك قواعد البيانات لا يقتصر على تلقي الاستفسارات البحثية من مستخدمي محرك البحث، وإنما يمتد إلى التفاعل والتعامل مع نماذج

بحث قواعد بيانات الويب ؛ حيث يقوم النموذج العام الموجود في واجهة بحث البحث بإرسال استفسار البحث إلى قاعدة أو قواعد بيانات الويب المتوافقة مع موضوع الاستفسار. ومن ثم فإن ثمة رابط لا ينقطع يجمع بين نموذج البحث في المحرك وبين نماذج البحث في قواعد البيانات، يقوى هذا الرابط أو يضعف بحسب مدى التوافق في البنية وآليات العمل بين نموذج البحث في المحرك ونماذج البحث في قواعد البيانات، فإذا ما كان التوافق في أعلى درجاته كان التفاعل بين المحرك وبين قواعد بيانات الويب في أفضل حالاته، ومن ثم تبلغ نسبة الدقة في النتائج المسترجعة أفضل مقاييسها، والعكس من ذلك صحيح.

الجدول (٥-٦) نموذج البحث في محركات بحث قواعد بيانات الويب غير المرئية

نموذج البحث		محركات بحث الويب غير المرئية					
		business	incywincy	infomine	scirus	scienceresearch	turbo10
نموذج عام		√	---	---	---	√	√
نموذج متعدد	فراغ النص	---	√	√	√	---	---
	زر التحديد	---	√	---	---	---	---
	زر الخصائص	√	---	√	√	√	---
	قائمة الاختيارات	---	√	√	√	---	---
	عنصر الترتيب والعرض	---	√	√	---	---	---

يتضح من الجدول (٥-٦) مجموعة من الملاحظات هي:

أولاً: انعكس واقع واجهات البحث في محركات عينة الدراسة على حال تواجد نماذج البحث ؛ وذلك حيث يكاد ينتفي وجود نموذج البحث في واجهات البحث البسيطة أو المباشرة، بينما يوجد نموذج البحث مصاحبًا لمجموعة من آليات البحث في واجهات البحث المتعددة أو المركبة. وعلى ذلك فإن بنية نموذج البحث هي جزء أساسي من البنية العامة التي ستكون عليها واجهة بحث المحرك.

ثانيًا: تعطي نماذج البحث المتعددة فرصًا أكبر للمستفيد في تنقيح وتدقيق استراتيجيات البحث ؛ حيث تساعد نماذج البحث ذات العدد الوفير من عناصر وحقول البحث المستفيد في اختيار ما يلائمه من خيارات البحث التي تعبر بدقة عن موضوع بحثه والتعريف بجوانبه المختلفة. كما أن نماذج البحث المتعددة لا تحوي فقط خيارات وصف الموضوع أو المعلومات المطلوبة، وإنما قد تحوي أيضًا خيارات وصف المواد أو وسائط المعلومات التي تلبي الحاجة الموضوعية، مثل خيارات حجم الملفات المطلوبة وشكل المعلومات.

ثالثًا: تعكس نماذج البحث المستخدمة داخل محركات بحث عينة الدراسة واقع تنظيم المعلومات داخل المحركات، وطبيعة تنظيم المعلومات داخل مواقع الويب غير المرئية ؛ فعندما يحتوي نموذج البحث المتعدد على خيارات البحث بالحقول الببليوجرافية، فإن ذلك يعني أن محرك البحث هذا يتعامل مع قواعد بيانات ذات بنية الجداول والتسجيلات، إلى جانب احتوائه على تسجيلات مكونة من حقول ببليوجرافية قابلة للبحث بواسطة المستفيد. على الجانب الآخر فإن عدم وجود خيارات البحث بالحقول الببليوجرافية يعني أن محرك البحث هذا يتعامل مع مواقع الويب التي لا تحوي تسجيلات ببليوجرافية أو تستخدم نظامًا مغايرًا لنظام الجداول والتسجيلات، وبذلك فهو يرتبط بالمواقع أكثر من ارتباطه بقواعد البيانات.

رابعًا: إن أفضل نماذج البحث المتعددة هو ما جاء في محرك البحث incywincy، وذلك ليس لأنه يحتوي على أكبر عدد من خيارات وعناصر البحث، وإنما لاحتوائه على خيارات وعناصر البحث الأكثر ملاءمة للتعامل مع مواقع وقواعد بيانات الويب غير المرئية، فقد تعددت خيارات البحث التي ذكرت فيها خصائص نماذج بحث مواقع وقواعد بيانات الويب.

١/٣/١/٢/٥ خيارات استرجاع نماذج بحث قواعد بيانات الويب:

يقصد بخيارات استرجاع نماذج بحث مواقع وقواعد البيانات هنا أن يوفر محرك بحث الويب غير المرئية بعض خيارات البحث التي تستطيع بحث واسترجاع واجهات بحث قواعد بيانات الويب. والفارق في هذه الحالة أن الهدف من عملية البحث والاسترجاع ليست صفحات أو تسجيلات قواعد البيانات، وإنما الهدف هو نموذج بحث قاعدة البيانات ذاته، ويتم استرجاع هذه النماذج لأهداف عدة منها أن مستخدم محرك بحث الويب غير المرئية قد يجد بين النماذج المسترجعة ما يلائمه لبحث الويب والحصول على المعلومات المناسبة، وبذلك يكون محرك بحث الويب غير المرئية قد فتح نوافذ للبحث أمام المستفيد بعدد ما تم استرجاعه من نماذج بحث مواقع وقواعد بيانات الويب. ويوضح الجدول (٥-٧) واقع خيارات بحث النماذج داخل محركات بحث عينة الدراسة.

الجدول (٥-٧) بحث نماذج بحث قواعد البيانات

بحث نماذج بحث قواعد البيانات	محركات بحث الويب غير المرئية					
	business	incywincy	infomine	scirus	scienceresearch	turbo10
يهتم باسترجاع المحتوى	√	√	√	√	√	√
يهتم باسترجاع نماذج البحث	---	√	---	---	---	---

يوضح الجدول (٥-٧) ما يلي:

إن أغلب محركات بحث عينة الدراسة المختصة ببحث الويب غير المرئية تملك نموذجًا للبحث يحوي في تركيبه عناصرَ وخيارات لبحث واسترجاع المعلومات غير المرئية داخل قواعد بيانات الويب. وقد تفرد محرك incywincy عنها جميعًا في امتلاكه لنموذج بحث يحوي خيارات مميزة تستطيع استرجاع نماذج بحث مواقع وقواعد بيانات الويب، دون استرجاع المحتوى الداخلي لها، أي أن مستوى

الاسترجاع هنا يقف عند حد استرجاع واجهات ونماذج البحث لمواقع الويب غير المرئيـة، دون أن يتعـدى ذلـك إلى مـستوى اسـترجاع المحتـوى والتـسجيلات الببليوجرافية لقواعد البيانات. وهو ما سيجعل من محركات بحث الويب غير المرئية التي ستتبع هذا الأسلوب بوابات عملاقة لبحث محتوى الويب المرئية وغير المرئية.

٤/١/٢/٥ آليات البحث المستخدمة:

تمثل آليات البحث المكون الثالث من مكونات محركات البحث بجانب برنامجي الزاحف والمكشف. وترتبط آليات البحث باعتماد محرك البحث على أسلوب البحث الحر بالكلمات الدالة، ومن ثم تأتي آليات البحث لتساعد على تكوين الشكل الملائم للبحث بالكلمات البحثية. وتتعدد أنواع آليات البحث بين تلك التي تستخدم مع الكلمات المفردة مثل ؛ البحث بحساسية الحروف، وآلية البحث الحر wild card، وتلك الآليات التي تستخدم مع الكلمات المتعددة مثل البحث بالتطابق، والبحث بالعبارات واستخدام المنطق البوليني boloean logic. وإذا كانت آليات البحث قد أخذت أهمية في عمل أدوات بحث الويب المرئية، فإنها ستأخذ أهمية أكبر في أدوات بحث الويب غير المرئية ؛ ذلك لأن دور آليات البحث هنا يكمن في التعامل مع نماذج بحث قواعد بيانات الويب التي تملك أشكالاً مختلفة من آليات البحث، ومن ثم وجب التوافق بين الآليات التي تعتمد عليها محركات بحث الويب غير المرئية، وبين تلك المستخدمة في قواعد بيانات الويب.

تساعد آليات البحث في إمكانية التحكم في صنع أشكال مختلفة من البحث بمجموعة من الكلمات أو بكلمات مفردة. وتختلف استراتيجيات البحث بدورها وطبيعتها عن آليات البحث التي تتكامل لتحقيق أنواع أشكال مختلفة من الاستراتيجيات. وتتعدد استراتيجيات البحث المستخدمة في محركات البحث

مثل ؛ البحث الحر بالكلمات المفتاحية واستخدام الحقول في بحث مصادر وصفحات الويب، ويمكن القول إن الويب يتم بحثها من خلال مجموعة مختلفة من استراتيجيات البحث مثل استخدام البحث بالمسار، أو استخدام الأدلة الموضوعية، أو استخدام محركات وكشافات البحث على اختلاف أنواعها المتخصصة والعامة أو المتعددة التي تبحث في محركات أخرى Meta search Engines أو استخدام قواعد بيانات الويب. بينما تهتم آليات البحث بتكوين استفسار البحث بطرق مختلفة مثل البحث بالتطابق وحساسية الحروف. ويوضح الجدول (٥-٨) آليات البحث المستخدمة في محركات بحث العينة، وهي :

الجدول (٥-٨) آليات البحث المستخدمة في محركات بحث عينة الدراسة

آليات البحث المستخدمة	محركات بحث الويب غير المرئية					
	business	incywincy	infomine	scirus	scienceresearch	turbo10
البحث بالتطابق	√	√	√	√	√	---
البحث الحر ? , *	❖ في أخر الكلمة	❖ في أخر الكلمة	❖ في أخر الكلمة	√	❖ في أخر الكلمة	❖ في أخر الكلمة
حساسية الحروف	---	---	---	---	---	---
بحث بالعبارات والتوقف	---	√	---	---	---	---
معاملات المنطق البوليني	√	√	√	√	√	and
البحث برابط التسجيلة	---	---	√	√	---	---
التصحيح الإملائي	---	√	√	√	---	√

يشير الجدول (٥-٨) إلى مجموعة من الملاحظات هي:

أولاً: إن اختبار وجود آليات البحث ضمن محركات عينة الدراسة ليس من قبيل استكمال شكل نموذج البحث داخل المحركات، وإنما لأجل تحقيق الدقة والكفاءة الداعمة للمستفيد في صياغة وتعديل استفسارات البحث كما يتناسب وموضوع الاستفسار. ومن ثم فإن محرك الويب غير المرئية الذي يملك القدر

الأكثر من الآليات في نموذج بحثه، يكون هو الأكفأ والأقدر على استرجاع محتوى الويب غير المرئية، بل والوصول إلى أعلى درجات الدقة والتحقيق precision.

ثانيًا: امتلكت محركات incywincy, infomine and scirus أفضل وأكثر الآليات التي تحويها نماذج البحث بالكلمات المفتاحية. ويرجع هذا إلى تركيز هذه المحركات الثلاث على أسلوب البحث الحر بالكلمات المفتاحية وإعطائه الأولوية في استرجاع المحتوى غير المرئي. بينما كانت المحركات الثلاث الأخرى في درجة متأخرة عن ذلك ؛ حيث إنها لم تحمل الكثير من آليات بحث الكلمات المفتاحية، ويعزو الباحث ذلك إلى أن محركات turbo10, sciencesearch, buseiess ذات إمكانات وتقنيات بحث محدودة تجعلها مجرد أدوات حصر وترتيب للويب غير المرئية، ولا تمكنها من أن تكون محركات بحث ذات كفاءة عالية، هذا فضلاً عن أن محرك البحث turbo10 إنما يعتبر أحد محركات البحث المتعددة، ولا يملك من صفحة البحث عن المحتوى غير المرئي سوى المسمى.

ثالثًا: استخدمت آليات بحث الكلمات المفتاحية بدرجات مختلفة داخل محركات عينة الدراسة ؛ فقد تواجدت آليتا البحث باستخدام المنطق البوليني والبحث الحر wild card في كل محركات العينة، وجاءت بعدهما آلية البحث بالتطابق التي تواجدت في كل محركات العينة ماعدا محرك turbo10. وانتفى تمامًا استخدام آلية البحث بحساسية الحروف في أي من محركات بحث العينة. ويرى الباحث أن واقع تواجد آليات البحث في محركات العينة يعكس طبيعة استخدام كل منها ؛ حيث تصل آلية البحث بالمنطق البوليني ومعاملاتها and, or, not من الأهمية للدرجة التي تجعل محركات بحث الويب على اختلاف أنواعها تستخدمها بشكل ذاتي، حتى وإن لم يرد ذكر هذه الآلية في جوانب عمل محرك البحث. بينما لم يستخدم أي من محركات بحث العينة آلية البحث

بحساسية الحروف، وذلك إنما يأتي من أن المستفيد سيكون على علم تام بما يضعه في استفساره من كلمات البحث ذات المفهوم والشكل المحددين مسبقًا، مما لا يجعل محرك البحث يهتم بشكل أو طريقة كتابة معينة، لافتراضه أن المستفيد على دراية تامة بما يصيوغه من كلمات بحث. وينطبق هذا الكلام أيضًا على آلية البحث بالعبارات والتوقف التي يعكس عدم تواجدها في أي من محركات البحث، افتراض أن مختلف الكلمات التي يضعها المستفيد داخل فراغ البحث، سواء أكانت كلمات ذات دلالة أم حروف وأدوات، فإنما تساهم جميعًا في التعبير عن جوانب موضوع البحث المختلفة.

٢/١/٢/٥حقول البحث الببليوجرافية:

يقصد بالبحث بالحقول هنا الاعتماد على رموز حقول الوصف الببليوجرافي في استرجاع وحدات من المعلومات تشترك معًا في مجموعة من الحقول مثل تاريخ نشر محدد أو مؤلف واحد أو موضوع مخصص. تختلف محركات بحث الويب غير المرئية عن محركات بحث الويب المرئية في أن الثانية تمتلك قاعدة بيانات تحتوى بداخلها على مجموعات المعلومات التي يملكها محرك البحث، وهذه المعلومات قد جاءت إليه نتيجة لعمل برنامج الزاحف في نسخ وإضافة مواقع وصفحات الويب إلى قاعدة بيانات المحرك. غير أن محرك بحث الويب غير المرئية سيتوقف هنا عند حد الوسيط في البحث بين المستفيد وبين قاعدة البيانات، فضلاً عن أن هذا المحرك لن يملك القدرة على احتواء مختلف معلومات قواعد بيانات الويب غير المرئية التي سيصل إليها، إنما الدور المنوط به فقط هو إجراء البحث في أكثر من قاعدة بيانات في آن واحد، تمامًا مثل ما يقوم به محرك البحث المتعدد عند إجراء البحث في محركات بحث الويب المرئية. ومن ثم فإن احتواء نموذج البحث في محركات الويب غير المرئية على خيارات البحث بالحقول الببليوجرافية، إنما هو للتوافق مع قواعد بيانات الويب، وليس لبحث مجموعاتها الداخلية.

إن أهمية خيار البحث بالحقول الببليوجرافية داخل نموذج بحث محرك الويب غير المرئية لا تكمن فقط في وجوده ضمن خيارات البحث الأخرى، وإنما الأهم من ذلك يأتي في حسن صياغة الحقول ودقة بنائها بما يتوافق والبحث في قواعد بيانات متعددة الأحجام من المعلومات، ومختلفة في وسائط المعلومات التي تحويها، ومتنوعة في نظم إدارتها وتشغيلها. ولذلك فإن الباحث يرى أن وجود الحقول الببليوجرافية ضمن خيارات نماذج البحث في محركات البحث غير المرئية يحتاج إلى البنية العامة الشاملة في مسميات الحقول الببليوجرافية ؛ فعلى سبيل المثال يمكن لمسمى (المصدر) أن يتوافق والمسميات الأخرى المختلفة المترادفة له التي تستخدمها قواعد بيانات الويب مثل (المؤلف، المنتج، والمنشئ). واعتماداً على ذلك يستطيع محرك بحث الويب غير المرئية امتلاك نموذجا عاما لبحث قواعد البيانات يمكن للمستفيد من خلاله تشكيل استراتيجية البحث وتضمين كل جوانب الموضوع.

الجدول (٥-٩) البحث بحقول البحث الببليوجرافية في محركات بحث عينة الدراسة

حقول البحث الببليوجرافية	محركات بحث الويب غير المرئية					
	business	incywincy	infomine	scirus	scienceresearch	turbo10
المؤلف أو المنشئ	---	---	√	√	---	---
العنوان	---	---	√	√	---	---
رؤوس الموضوعات	---	---	√	√	---	---
الجهة أو الناشر	---	---	√	√	---	---
النوع أو الشكل	---	---	√	√	---	---
رابط ومحدد التسجيلة ID	---	---	√	√	---	---
تواريخ مرتبطة	---	---	√	√	---	---

يتضح من الجدول (٥-٩) مجموعة من الملاحظات هي:

أولاً: عمل محركا البحث infomine, scirus فقط من بين باقي محركات عينة الدراسة على تبني آلية البحث بالحقول الببليوجرافية، وذلك حتى يتمكن مستخدمو هذين المحركين من استرجاع التسجيلات الببليوجرافية النهائية التي تشكل محتوى قواعد بيانات الويب. ويعكس ذلك أن هذين المحركين هما الأكثر تفاعلاً مع محتوى قواعد البيانات تنظيمًا واسترجاعًا، خلاف باقي المحركات التي تتجه إلى مواقع الويب غير المرئية أكثر من اتجاهها إلى قواعد البيانات.

ثانيًا: إن ثمة رابط بين مجال التخصص لمحرك البحث وبين طبيعة صياغة نماذج البحث به ؛ حيث أثر مجال التخصص العلمي لمحركي البحث infomine, scirus على أن يتجها إلى التعامل مع قواعد بيانات الويب العلمية المتخصصة، ومن ثم فإن هذين المحركين احتاجا إلى بناء آليات البحث بالحقول الببليوجرافية التي تتميز عن غيرها من آليات البحث الأخرى، بالقدرة على استرجاع المحتوى غير المرئية لقواعد بيانات الويب.

ثالثًا: لم يكتف محركا البحث infomine, scirus بتوفير إمكانية البحث بالحقول الببليوجرافية المتعارف عليها ضمن تسجيلات الوصف الفني التي تخص الشكل أو الموضوع فقط، وإنما عمل هذان المحركان على توفير خيارات البحث بحقول ببليوجرافية أخرى تخص طبيعة الملفات التي تحوي المعلومات النهائية ؛ مثل ما جاء في محرك scirus من خيارات ؛ كخيار البحث بشكل الملف doc, pdf, html، وخيار البحث بطبيعة المعلومات النهائية إذا كانت مستخلص أو نص كامل.

٢/٢/٥ برامج زحف محركات بحث الويب غير المرئية:

يقوم برنامج الزاحف في محرك بحث الويب المرئية بدور التزويد والإضافة لصفحات ومواقع الويب، ويتم ذلك وفقًا لآليات عمل معدة مسبقًا تكمن في ملف

robot.txt الذي يرشد برنامج الزاحف لما سيقوم به وما يخص عمليات التحديث والزيارات الدورية لمواقع الويب المرئية. وكما يشكل برنامج الزاحف أول المكونات الثلاث الأساسية في محرك بحث الويب المرئية، فإن دوره في محرك بحث الويب غير المرئية لا يقل أهمية عن ذلك، غير أن برنامج الزاحف داخل محرك الويب غير المرئية لن يعمل بآليات متوافقة وبنية html فقط كما في زاحف الويب المرئية، بل يتركز الأمر على التعامل مع واجهات ونماذج بحث الويب غير المرئية الممثلة في نماذج واجهات بحث قواعد بيانات الويب، ونماذج بحث مواقع الويب غير المرئية.

وعلى ذلك فإن برنامج الزاحف في الويب غير المرئية سوف يعمل على اكتشاف وتحليل نماذج البحث search forms بدلاً من اكتشاف وتعقب الروابط الفائقة URLs. وللقيام بمهام تكشيف وتحليل مواقع وقواعد بيانات الويب غير المرئية، يجب على برنامج الزاحف التحلي بمجموعة من آليات العمل التي تختلف كلية عما كان عليه داخل محركات الويب المرئية. ومن هذه الآليات ؛ إمكانية التعرف إلى الواجهات التي تحوي نماذج البحث داخل مواقع وقواعد الويب غير المرئية، وتمييزها عن غيرها من مواقع الويب الأخرى، أيضًا التحلي بآلية القدرة على تحليل وتكشيف هذه النماذج لأجل تضمينها ضمن مجموعات محرك الويب غير المرئية. وسوف يحاول الباحث استقراء واقع برنامج الزاحف داخل محركات بحث عينة الدراسة، فضلاً عن خصائص عمله في حال نشاطه داخل محرك البحث.

إن محركات بحث الويب المرئية أو غير المرئية ترتبط ارتباطًا وثيقًا بواقع برنامج الزاحف بها ؛ حيث يمثل تواجد برنامج الزاحف وجه الاختلاف الأساسي بين المحرك وبين الدليل كأداتين لبحث الويب. ومتى كان برنامج الزاحف متوقفًا عن العمل أو لا يوجد ضمن مكونات محرك البحث، فإن محرك البحث هذا لا يعد محرك بحث للويب من الناحية الدلالية والتطبيقية. وهذا الوصف

ينسحب أيضًا على محركات بحث الويب غير المرئية، وليس فقط محركات الويب المرئية. وسوف يتم قراءة هذا الوصف من واقع عينة محركات بحث الويب غير المرئية في الجدول (٥-١٠) كما يلي :

الجدول (٥-١٠) برامج الزحف داخل محركات بحث الويب غير المرئية عينة الدراسة

برامج زحف محركات البحث		محركات بحث الويب غير المرئية					
		business	incywincy	infomine	scirus	scienceresearch	turbo10
فاعلية الزاحف	نشـط	---	√	---	√	---	---
	غير نشط	√	---	√	---	√	√
اتجاه العمل بروابط الويب	للصفحات	---	√	---	√	---	---
	للنماذج	---	√	---	---	---	√
تعامل المحرك مع قواعد البيانات	تكشيف عام	√	√	√	√	√	√
	تكشيف المحتوى	---	√	---	√	---	√
تكشيف نماذج البحث		---	√	---	---	---	√

يتضح من الجدول (٥-١٠) مجموعة من الملاحظات يتم تناولها في الجوانب التالية:

١/٢/٢/٥ تواجد الزاحف في محركات بحث الويب غير المرئية:

أولاً: لقد كشف غياب برنامج الزاحف عن بعض محركات بحث عينة الدراسة حقيقة هذه المحركات في أنها لا تنتمي إلى محركات بحث الويب غير المرئية، وإنما يمكن تصنيفها على أنها أنواع أخرى من أدوات البحث ونظم الاسترجاع التي تحاول العمل على استرجاع المحتوى غير المرئي ؛ لأن المحركات التي لا تملك برنامج الزاحف لا تستطيع أن تجوب الويب باحثة عن نماذج وصفحات بحث قواعد البيانات أو مواقع الويب غير المرئية. وقد وُصفت بعض

محركات بحث عينة الدراسة بوصف آخر غير كونها محرك بحث للويب غير المرئية ؛ حيث ذكر محرك البحث infomine أنه مكتبة تخيلية virtual library للمصادر الرقمية على الويب، كما وُصف محركا البحث business, sciencesearch عملهما بأنهما بوابات portal لمصادر المعلومات العلمية وإدارة الأعمال.

ثانيًا: لقد وجد برنامج الزاحف عمليًا في اثنين فقط من محركات بحث عينة الدراسة، وعلى الرغم من ذلك فإنهما لا يتساويان في عملهما ؛ حيث يمثل محرك incywincy نموذجًا فعليًا لمحرك بحث الويب غير المرئية دلالةً وأداءً، أما محرك البحث scirus فهو بالأساس محرك بحث للويب المرئية، غير أنه يعمل على الوصول إلى قواعد بيانات الدوريات العلمية وتكشيفها لاسترجاع محتوياتها من المقالات والنصوص الكاملة.

ثالثًا: يعتمد محرك البحث المتعدد turbo10 على آلية عمل form finder التي تشبه عمل برنامج الزاحف في محرك البحث؛ حيث إن هذه الآلية ليست برنامجًا مصممًا لزحف الويب واكتشاف صفحات كما يعمل برنامج الزاحف، وإنما تعمل فقط على الوصول إلى نماذج البحث في واجهات محركات البحث وتحليلها لضمها ضمن قائمة محركات البحث التي يبحث داخلها محرك turbo10. ومن ثم تعمل هذه الآلية على تحسين أداء turbo10 كمحرك بحث متعدد.

٢/٢/٢/٥ اتجاه مسح روابط الويب:

إن برنامج الزاحف في محرك بحث الويب غير المرئية المقترح سيحمل عبء الوصول إلى نماذج البحث في واجهات مواقع وقواعد بيانات الويب غير المرئية، وذلك لتحليلها وتكشيفها، ومن ثم فإن اتجاهه في العمل سيقوم على تتبع نماذج البحث وليس صفحات الويب ذات بنية html.

لقد ركز اثنان فقط من محركات بحث العينة هما incywincy, turbo10 على الوصول إلى نماذج بحث مواقع الويب، غير أن كلاً منهما يحمل سببًا

مختلفاً لذلك ؛ فمحرك البحث incywincy يسعى إلى الوصول لنماذج بحث مواقع وقواعد بيانات الويب غير المرئية حتى يكشفها ويخضع محتواها للبحث من جانب مستخدمي الويب، أما محرك البحث turbo10 فيعمل على الوصول إلى نماذج بحث محركات بحث الويب حتى يقوم بتحليلها والتعرف إلى خصائص البحث بها، ومن ثم ضمها إلى قائمته من محركات البحث التي يبحث داخلها.

٣/٢/٢/٥ تكشيف وبحث قواعد البيانات:

لقد انتهجت محركات بحث العينة طريقتين في التعامل مع قواعد بيانات الويب تكشيفاً وتحليلاً؛ فالطريقة الأولى تقف عند حد الوصف الموضوعي العام لمجال أو تخصص قاعدة البيانات، وذلك بأن يتم التعبير عن مجال قاعدة بيانات الويب برأس موضوع أو مصطلح أو أكثر للدلالة على محتواها، أما الطريقة الثانية فتعمل على تحليل وتكشيف المحتوى الداخلي لقاعدة البيانات، والتعرف إلى موضوع ومحتوى كل تسجيلة أو وحدة من مجموعاتها الداخلية، وذلك بهدف بحث واسترجاع محتواها عند توجيه استفسار البحث إليها.

عملت محركات بحث العينة incywincy, scirus, turbo10 على الوصول إلى مجموعات قواعد بيانات الويب وتحليلها، غير أن محرك البحث turbo10 لا يهدف إلى تحليل قواعد بيانات الويب مثل محرك incywincy, scirus، وإنما يهدف للوصول إلى قواعد بيانات محركات بحث الويب المرئية تحديداً. على الجانب الآخر فإن محركات business, sciencesearch, infomine عملت على التحليل العام لقواعد بيانات الويب، وذلك لأجل ترتيبها في قوائم حصرية أو تقسيمات موضوعية، وهو ما يتلاءم وطبيعة هذه الأدوات في كونها تميل إلى عمل أدلة البحث أكثر من عملها كمحركات لبحث الويب غير المرئية، في حين أن هذه الأدوات تملك إمكانية توجيه استفسارات البحث إلى قواعد البيانات واسترجاع محتواها.

٤/٢/٢/٥ تكشيف نماذج بحث قواعد البيانات:

إن وصول برنامج الزاحف إلى نماذج بحث مواقع وقواعد بيانات الويب غير المرئية ليس هو الغاية في حد ذاته، وإنما تكمن الغاية في إمكانية تحليل وتكشيف هذه النماذج للتعرف إلى خصائصها ومكوناتها وعناصر البحث التي تتركب منها. وهذا كله يهدف بدوره إلى فهم هذه النماذج من حيث آليات البحث بها وطبيعة عملها، حتى تتمكن محركات بحث الويب غير المرئية من توجيه استفسارات البحث إليها، واسترجاع مجموعات قواعد البيانات بواسطتها.

٣/٢/٥ واجهات عرض النتائج المستدعاة من قواعد البيانات:

تولي أدوات بحث الويب جميعها أهمية خاصة لواجهات عرض نتائج البحث ؛ حيث تمثل هذه الواجهات ثمرة الرد على استفسار البحث الخاص بالمستفيد. ولا تقتصر عناية أدوات بحث الويب بواجهات عرض النتائج على الشكل العام لواجهة العرض، وإنما يكون الاهتمام الأكبر بطرائق ترتيب وتنظيم الوحدات المسترجعة بحسب درجة صلتها بموضوع البحث. ويزيد هذا الأمر في أهميته عندما تسترجع أداة البحث على الويب مجموعات مختلفة من نتائج البحث يتم تلقيها من قواعد بيانات مختلفة، مما يعني أن تقوم أداة البحث هنا بعمليات الفرز والاستبعاد والتنظيم وإعادة الترتيب مرة أخرى، وهو ما يحدث تمامًا في محركات البحث المتعددة، وما سيحدث في محركات بحث الويب غير المرئية المقترحة. ومن ثم فإن الاهتمام بآليات عرض النتائج لا يقل أهمية عن باقي العمليات التي يقوم بها محرك بحث الويب غير المرئية. ويوضح الجدول (٥-١١) طبيعة التعامل مع النتائج المستدعاة داخل واجهات عرض نتائج محركات بحث عينة الدراسة.

الجدول (٥-١١) واجهات عرض النتائج المستدعاة لمحركات بحث الويب غير المرئية عينة الدراسة

واجهات عرض النتائج المستدعاة		محركات بحث الويب غير المرئية					
		business	incywincy	infomine	scirus	scienceresearch	turbo10
خيارات عرض وترتيب النتائج	أثناء الاستفسار	---	عدد الوحدات	√	---	---	---
	أثناء العرض	√	---	---	√	√	---
خصائص الفرز والاستبعاد	ذاتية	√	---	√	√	√	√
	اختيارية	---	√	---	---	---	---
أنماط ترتيب النتائج	بالموضوع	√	√	√	√	√	√
	الحداثة	---	---	---	√	---	---
	أخرى	مصدر الوحدة	---	عنوان الوحدة	الروابط	مصدر الوحدة	---

يشير الجدول (٥-١١) إلى مجموعة من الملاحظات يتم تناولها في الجوانب التالية:

١/٣/٢/٥ خيارات ترتيب نتائج البحث:

عملت محركات بحث عينة الدراسة جميعها على توفير خيارات عرض وترتيب نتائج البحث المسترجعة أمام المستفيد، وقد شذ عن ذلك فقط محرك البحث trubo10 الذي لم يوفر أية خيارات تخص التحكم في عرض النتائج المسترجعة سواء أثناء البحث أم أثناء عرض النتائج المستدعاة. على الجانب الآخر فقد تباينت محركات بحث عينة الدراسة في طبيعة وكم خيارات التعامل مع النتائج المسترجعة ؛ حيث اكتفى بعض محركات البحث بتقديم خيار التحكم في عدد الوحدات المسترجعة المعروضة في صفحة، وتعدى بعضها الآخر ذلك إلى تقديم خيارات تخص أنماط ترتيب النتائج سواء بالموضوع أو تبعًا لحداثة المعلومات. وقد

حرصت بعض محركات بحث العينة مثل infomine على تقديم خيارات متقدمة لعرض النتائج تتمثل في شكل العرض ؛ حيث يمكن عرض النتائج المسترجعة في شكل عناوين فقط أو في شكل مختصر أو في شكل تسجيلة كاملة.

٢/٣/٢/٥ خصائص الفرز والاستبعاد:

تستقبل محركات البحث المتعددة ومحركات بحث الويب غير المرئية مجموعات متعددة من نتائج البحث المسترجعة من أكثر من قاعدة بيانات أو موقع ويب، ومن ثم فإن ملامح التكرار والتداخل وعدم الفرز سوف تصيب هذه المجموعات المسترجعة، لذا كان من الضروري أن توفر محركات بحث الويب غير المرئية آليات تعمل على تصفية وتنقية وفرز مجموعات النتائج المسترجعة، بحيث تمنع المتكرر وتعيد الترتيب مرة أخرى بما يتوافق وملاءمة العرض أمام المستفيد. وقد لاحظ الباحث أن جميع محركات بحث عينة الدراسة تقوم بعمليات الفرز والاستبعاد ذاتيًا قبل عرض النتائج أمام المستفيد، وهو ما يتم آليًا؛ لكونه أحد الخصائص الذاتية، غير أن محرك البحث incywincy قد رأى أمرًا مغايرًا يكمن في أن قرار الفرز وحجب التكرار يجب أن يكون بقرار من المستفيد ذاته، فقد يوجد من المستفيدين من يتطلع إلى رؤية النتائج المكررة حتى يبحثها جميعًا، فيقبل ما يقبله ويرفض ما يرفضه.

٣/٣/٢/٥ أنماط عرض الوحدات المسترجعة:

تذخر الويب المرئية وغير المرئية بملايين الصفحات والمواقع وقواعد البيانات، وعلى ذلك فإن البحث عن أدق الموضوعات وأخصها قد يؤدي إلى استرجاع عشرات الآلاف من نتائج البحث ووحدات المعلومات. ومن ثم فإن قدرة المستفيد على تصفح هذا الكم الهائل من النتائج تقل بازديادها العددي وتنوعها الشكلي. وقد عملت محركات الويب جميعها على اتباع آلية لترتيب النتائج المسترجعة تبعًا

لأهميتها ومدى صلتها بموضوع البحث، وتبعًا لهذه الآلية فإن أقرب النتائج لموضوع البحث هي التي تأتي في مقدمة النتائج المسترجعة، ومن ثم يمكن للمستفيد أن يقتصر في تصفحه على أولى النتائج أو أولى الصفحات في واجهات عرض النتائج. وتنتهج بعض محركات البحث سبلاً أخرى في ترتيب النتائج المسترجعة، خلاف الأقرب لموضوع البحث، من هذه السبل؛ الترتيب بحسب مدى حداثة المعلومات المسترجعة، أو الترتيب بحسب أهمية مصادر أو قواعد البيانات التي تم البحث فيها. وقد تبنت محركات البحث جميعها مبدأ الترتيب بحسب الصلة للموضوع، وجمع بعضها بين الترتيب بالموضوع وبين الترتيب بحداثة المعلومات كما في محرك البحث scirus، أو الترتيب بحسب مصدر النتائج، كما في محرك البحث business.

٤/٢/٥ خدمات محرك بحث الويب غير المرئية:

إن عمل محركات بحث الويب المرئية وغير المرئية لا يقف فقط عند توفير نماذج وآليات البحث واسترجاع النتائج وترتيبها، بل يتعدى ذلك إلى تقديم بعض الخدمات الأخرى التي تسهل من عملية البحث والتحكم في مجموعات النتائج المسترجعة ؛ مثل البحث داخل النتائج أو إعادة ترتيب النتائج بشكل أو بآخر أو بناء حساب خاص للمستفيد داخل محرك البحث يمكنه من حفظ نتائج البحث واستدعائها عند الحاجة إليها، بما يوفر معه إجراء عمليات البحث مرة أخرى. وقد اهتم الباحث بالحديث عن هذه الخدمات لأنها جانب مهم لا يمكن إهماله في عمل مختلف أدوات البحث التي تسعى لتقديم خدمات بحث نافعة وقوية، وإنما تسعى لتسهيل وتوفير وقت وجهد المستفيدين بأشكال مختلفة. ويقدم الجدول (٥-١٢) بعض نماذج الخدمات التي يمكن أن تقدمها محركات بحث الويب غير المرئية، وواقع هذه الخدمات داخل عينة الدراسة.

الجدول (٥-١٢) خدمات محرك بحث الويب غير المرئية عينة الدراسة

خدمات محرك بحث الويب غير المرئية	محركات بحث الويب غير المرئية					
	business	incywincy	Infomine	scirus	scienceresearch	turbo10
تحديد اتجاهات البحث	√	---	√	√	√	---
البحث داخل النتائج	√	---	√	√	√	---
التواصل مع محرك البحث	√	√	√	√	√	√
البريد الإلكتروني	---	√	---	√	---	---
إحصاءات حول النتائج	---	---	√	√	---	√
بناء سمات المستفيد	---	√	---	---	---	---
التوزيع الموضوعي للنتائج	---	√	---	---	---	√

يتضح من الجدول (٥-١٢) النتائج التالية:

١/٤/٢/٥ تحديد اتجاهات البحث:

تتمثل خدمة تحديد اتجاهات البحث في منح محرك بحث الويب غير المرئية للمستفيد خيارات تحديد قاعدة أو قواعد البيانات التي سيقوم بالبحث داخلها فقط. وتأتي أهمية هذه الخدمة في أنها تتيح للمستفيد اختيار قواعد البيانات التي يراها أقرب للإجابة عن استفساره وأكثر تخصصًا في موضوع البحث. ولكي يتمكن محرك بحث الويب غير المرئية من تقديم هذه الخدمة، لابد أن يعمل على حصر وترتيب قواعد البيانات التي يتعامل معها في نظام محدد يمكن المستفيد من تصفح هذه القواعد، واختيار ما يجده ملائمًا للبحث داخله. وقد أولت محركات بحث عينة الدراسة أهمية كبيرة لهذه الخدمة، ما جعل أربعة منها تتيح هذه الخدمة أمام المستفيدين.

٢/٤/٢/٥ البحث داخل النتائج:

تتيح خدمة البحث داخل النتائج للمستفيد تضييق جوانب الاستفسار الموضوعية؛ حيث يمكن للمستفيد تصفح النتائج المسترجعة عند طرحه

لاستفسار البحث في بداية الأمر ، وإذا وجد المستفيد النتائج المسترجعة غير مرضية أو أنها تحتاج إلى إعادة ترتيبها ، فإن خدمة البحث داخل النتائج تتيح له أن يبحث داخل النتائج المسترجعة فقط، ومن ثم يمكنه تضييق حدود الموضوع وحدود البحث على مجموعات النتائج المعروضة أمامه على واجهة محرك البحث. وتقدم محركات بحث الويب المرئية وغير المرئية هذه الخدمة تحت مسميات مختلفة مثل تنقيح البحث refine your search أو تحديده.

٣/٤/٢/٥ التواصل مع محرك البحث:

يحتاج محرك بحث الويب كغيره من نظم الاسترجاع إلى التغذية المرتدة أو التلقيم المرتد feed back ، وذلك لكي يتعرف محرك البحث إلى خصائص أدائه الإيجابية أو السلبية ، وتمثل آلية الاتصال أو التواصل مع المستفيد أهم سبل تفعيل التلقيم المرتد داخل محرك البحث ؛ حيث أجمعت محركات عينة الدراسة على التواصل مع المستفيدين والتعرف إلى آرائهم وردود أفعالهم من خلال خدمات contact us or E-mail us ، ومن ثم يمكن لمحرك البحث استقبال المعلومات المختلفة من جانب المستفيدين التي تعرف بخصائص العمل وجوانبه المختلفة.

٤/٤/٢/٥ البريد الإلكتروني:

تختلف خدمة البريد الإلكتروني عن الخدمة السابقة في أن هذه الخدمة لا تقتصر على التواصل مع محرك البحث فقط، وإنما شأنها هنا شأن باقي خدمات البريد الإلكتروني مثل hotmail, yahoo. وتقديم هذه الخدمة داخل محركات بحث الويب غير المرئية لا يتعلق بأي جانب من جوانب بحث المحتوى غير المرئي في شيء، وإنما هي واحدة من الخدمات التي تلائم عمل بوابات الويب web portal أكثر من كونها تلائم أدوات البحث. وقد اقتصرت هذه الخدمة على محركي البحث incywincy, scirus.

٥/٤/٢/٥ بناء سمات المستفيد:

انفرد محرك البحث incywincy عن باقي محركات بحث عينة الدراسة بتقديم خدمة تحديد ملامح أو سمات المستفيد، ويقصد بهذه الخدمة أن يتيح محرك البحث للمستفيد تحديد الخصائص والمتطلبات التي يحتاجها في بحثه، فضلاً عن أهم قواعد البيانات التي يرغب في البحث داخلها، ويمكن له تسجيل المصطلحات البحثية ذات الأهمية الخاصة به. ويرى الباحث أن هذه الخدمة هي الأكثر ارتباطًا بجوانب بحث الويب غير المرئية، وإذا ما أراد أي من محركات بحث الويب التزود بخدمات محددة، فإن هذه الخدمة ستكون الأنسب والأفضل بين مجموعة الخدمات.

٦/٤/٢/٥ التوزيع الموضوعي للنتائج:

عمل محركا البحث turbo10, incywincy على تقديم خدمة التوزيع الموضوعي لنتائج البحث المسترجعة أمام المستفيد ؛ ويقصد بهذه الخدمة أن يعمل محرك بحث الويب غير المرئية على تحليل وحدات النتائج المسترجعة والتعرف إلى موضوع كل وحدة منها، ومن ثم يستطيع محرك البحث حصر الموضوعات الناتجة عن عملية التحليل وتقدير عدد الوحدات التي تخص كل موضوع ثم تقديم قائمة حصرية بموضوعات النتائج المسترجعة مصاحبًا لعدد الوحدات. ومن خلال عرض قائمة الموضوعات يتمكن المستفيد من تحديد الجوانب الموضوعية للنتائج المستدعاة، بل ويستطيع التعرف إلى أغنى الجوانب الموضوعية من حيث عدد الوحدات المسترجعة، وقد يتيح محرك البحث للمستفيد أن يصل إلى مجموعة الوحدات الخاصة بكل جانب أو موضوع فرعي وعرضها في صفحة خاصة، وعلى ذلك يكون المستفيد قادرًا على التحرك داخل كم مجموعات النتائج في مسارات موضوعية مستقيمة.

٣/٥ أدلة بحث قواعد البيانات:

يطلق على هذا النوع من أدوات البحث مسميات كثيرة مثل الأدلة الموضوعية لمصادر الانترنت، أدلة الانترنت، كشافات الانترنت، فهارس الانترنت subject guides internet resource guides, internet directories, internet indices, internet catalogs وكلها تعطي معنى القوائم أو الموضوعات المقسمة. وتعتمد أدلة البحث على تقسيم المعرفة أو العلوم إلى أقسام عامة، ثم إلى أقسام أكثر تخصصًا أو قوائم فرعية، ومن ثم يمكن الحصول على الموضوع، ومن ثم المواقع التي ترتبط مباشرة بهذا الموضوع المخصص. وقد تعددت أنواع الأدلة على الويب، ولكن يمكن تقسيمها وفقًا للفئات التالية:

١. أدلة بحث موضوعية : وهي التي تعتمد على التقسيمات العلمية للموضوعات وهي بدورها تنقسم إلى نوعين أدلة البحث العامة، وأدلة البحث المتخصصة؛ أما العامة فهي التي تغطي مجالات المعرفة كافة، وهي قد لا تعتمد على تقسيم موضوعي محدد، وإنما تعتمد على تقسيم عام لكل المعرفة وقد يتضمن هذا التقسيم موضوعات خاصة بالأسرة أو الرحلات أو الترفيه. أو قد تعتمد على خطة تصنيف عالمية، أو خطة تصنيف قومية، أو متخصصة. أما أدلة البحث الموضوعية المتخصصة فقد تغطي موضوعات متخصصة مثل مجال المكتبات، أو الطب مثلاً.

٢. أدلة بحث نوعية : وهي التي توجه إلى فئة معينة من المستفيدين مثل أدلة مواقع الأطفال أو المسنين، أو ذوي الاحتياجات الخاصة، أو المرأة أو غيرها وهي بدورها قد تعتمد على خطة تصنيف عالمية، أو متخصصة.

٣. أدلة بحث جغرافية : وهي التي تعتمد على تغطية المواقع التي تنتمي إلى دولة معينة أو منطقة جغرافية تتحدث بلغة واحدة.

١/٣/٥ ملامح أدلة بحث قواعد البيانات:

تعمل أدلة بحث الويب بصفة عامة على مبدأ حصر وترتيب أو تقسيم مواقع ومصادر معلومات الويب، ومن ثم فإن أدلة بحث الويب المرئية وغير المرئية تعمل بآلية عمل وأسلوب واحد، وذلك خلاف محركات بحث الويب غير المرئية التي تحتاج إلى آليات عمل وخصائص مخالفة لتلك المتبعة في محركات بحث الويب المرئية. ويرى الباحث أن دليل البحث لا يعدو كونه أداة حصر وترتيب لمواقع الويب، ومن ثم فإن أهمية الحديث عنه هنا ترجع لكونه أسلوبًا مكملاً لأسلوب البحث الحر المتبع في محركات البحث، كما أنه يعد إحدى الخدمات البحثية التي يمكن لمحرك بحث قواعد بيانات الويب غير المرئية تقديمها للمستفيدين. وعلى ذلك فإن تناولنا لدليل بحث الويب غير المرئية سيركز فقط على جوانب التقسيم الموضوعي وطبيعة تكشيف وتحليل قواعد بيانات الويب وتقديمها في شكل قائمة أو دليل ذات تقسيم موضوعي منهجي. وسوف يتم تناول ملامح أدلة البحث من حيث واجهة دليل البحث، وطبيعة بناء أسلوب التصفح وخصائصه وأنماطه، وطبيعة تكشيف قواعد بيانات الويب التي سيتم إدراجها في قوائم دليل البحث.

١/١/٣/٥ واجهة دليل البحث:

تتسم واجهات بحث أدلة البحث دائمًا باحتوائها على التقسيمات أو القوائم الموضوعية التي تقسم مجموعاتها على أساسها، وهذا ما يعني أن عنصر البساطة غالبًا ما يغيب عن واجهات أدلة بحث الويب. ومن ثم فإن واجهة دليل البحث دائما ما تسعى إلى عرض الموضوعات أو الأقسام الموضوعية الأعم التي تعبر عن محتوياتها الداخلية، مما يعني أن الواجهة الرئيسة لدليل البحث هي الحلقة الأولى في السلسلة الموضوعية لأي قطاع موضوعي داخل دليل البحث ؛ حيث يبدأ المستفيد من القسم العام الأكبر في الموضوع، ثم الانتقال إلى الأقسام الأكثر تخصصًا حتى يرتضي المستفيد مستوىً محددًا من مستويات الموضوع لاسترجاع

المعلومات المتعلقة به. ويوضح الجدول (٥-١٣) أهم السمات التي قد تحويها واجهة دليل البحث، وهي كما يلي :

الجدول (٥-١٣) واجهة أدلة بحث الويب غير المرئية

واجهة أدلة البحث		أدلة بحث الويب غير المرئية					
		Academici-nfo	beaucoup	Complete-planet	Find-articles	genealogyspot	Mag-portal
شكل واجهة البحث	مفردة	√	√	- - -	√	- - -	√
	مركبة	- - -	- - -	√	- - -	√	- - -
أساليب البحث	البحث الحر	- - -	√	√	√	- - -	√
	بحث التصفح	√	√	√	للدوريات	عام جدًا	√

يتضح من الجدول (٥-١٣) النتائج التالية:

أولاً: عملت مختلف أدلة بحث العينة على توفير البحث داخل مجموعاتها اعتمادًا على أسلوب التصفح كأسلوب أساسي ثم الاستعانة بأسلوب البحث الحر كداعم للبحث السريع داخل قواعد البيانات. وقد شذ عن ذلك دليلا البحث genealogyspot, academicinfo ؛ حيث اعتمدا فقط على أسلوب التصفح في استرجاع محتويات مواقع وقواعد بيانات الويب غير المرئية التي تدعم تصفحها.

ثانيًا: جاءت واجهات أغلب أدلة البحث في العينة كواجهة مفردة غير مركبة، مما عكس واقع عمل هذه الأدلة في اعتمادها فقط على مستوى واحد أو اثنين على الأكثر في بناء قوائمها أو تقسيماتها الموضوعية.

٢/١/٣/٥ أساليب البحث بالتصفح:

"يمثل التصفح برؤوس الموضوعات أو المصطلحات داخل الويب نموذجًا من نماذج الاعتماد على اللغة المقيدة في التكشيف وتحليل محتوى مصادر الويب، لما

يرتبط به من تقييد مستخدم نظام الاسترجاع بقائمة مصطلحات هرمية التنظيم يتم ربطها بمجموعة مصادر المعلومات التي تتوافق مع دلالة كل من هذه المصطلحات. وهناك من أدوات بحث الويب المرئية مثل Yahoo ما يعتمد على قائمة من المصطلحات الخاصة بالموقع تندرج تحت منها صفحات الويب التي تلبي الحاجة الموضوعية الخاصة لهذه المصطلحات. ويمكن في هذه الحالة الاعتماد على أي من المكانز الخاصة بالتكشيف في مجال محدد أو مجالات متعددة بصياغة شكل من أشكال التصفح الداعم للتحكم في مجموعة ثابتة من المصطلحات تتكون على أثرها قوائم رئيسة ثم فرعية تدرج تحت منها محتويات قاعدة البيانات ككل. وقد اعتمدت بعض أدوات البحث على نظم تصنيف معدة مسبقا مثل تصنيف ديوي والتصنيف العشري العالمي، أو أن تعتمد بعض المواقع الأخرى على صياغة أشكال من نظم التصنيف تلائم المحتويات الخاصة بها من مصادر الويب. وقد ذكر هانئ عطية بعض جهود تصنيف المعلومات على الويب في سرد مجموعة من نظم التصنيف المعدة مسبقًا أو ذاتية البناء، بالإضافة إلى تقسيم تلك النظم إلى أنماط عامة متعددة مثل ؛ نظم التصنيف المكتبية كنظام مكتبة الكونجرس، ونظم التصنيف القومية كالتصنيف الأساسي في هولندا والتصنيف القومي السويدي للمكتبات العامة، ونظم التصنيف الموضوعية كالتصنيف الخاص بمجال الحاسب الآلي والتصنيف الموضوعي للرياضيات، ونظم التصنيف المولدة في البيئة الرقمية كتصنيف بوابة Edinburgh Engineering Virtual Library وتصنيف بوابة Social Science Information Gateway." وهذا يعني أن بعض أدوات البحث يعتمد في تنظيمه على قوائم رؤوس الموضوعات سواء أكانت معدة مسبقًا أو تم إعدادها خصيصًا لدليل البحث، فضلاً عن استخدام بعض أدلة البحث لأنظمة التصنيف المعيارية المختلفة[(١)].

(١) سيد ربيع سيد إبراهيم. محركات بحث الصور الثابتة على الإنترنت : دراسة تحليلية.- ط١.- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧م، ص ١١٤.

الجدول (٥-١٤) البحث بالتصفح داخل أدلة بحث الويب غير المرئية

البحث بالتصفح داخل دليل البحث		أدلة بحث الويب غير المرئية					
		Academici-nfo	beaucoup	Complete-planet	Find-articles	genealogyspot	Mag-portal
أنماط التقسيم الموضوعي	بالمصطلحات	√	√	√	√	جوانب الدولة ككل	√
	بالعناوين	√	√	√	√		√
	بالروابط	√	√	√	√	√	√
	بالنماذج المصورة	---	---	---	---	---	---
	أخرى	---	---	---	المقالات	---	---
أنظمة التقسيم الموضوعي	خطط تصنيف	---	---	---	---	---	---
	رؤوس موضوعات	---	---	---	---	---	---
	تقسيم عام	√	√	√	√	√	√

يتضح من الجدول (٥-١٤) النتائج التالية:

أولاً: أنماط التقسيمات الموضوعية:

تتبنى أدلة البحث على الويب نماذجَ مختلفة من أنماط تصفح مجموعاتها الداخلية، ونمط التصفح هنا يقصد به شكل التصفح والرابط الذي ينتقل منه المستفيد أثناء استعراضه للقوائم الموضوعية إلى استرجاع النتائج التي تتوافق وموضوع البحث ؛ فمن أدلة البحث ما يستخدم المصطلحات أو الكلمات، ومنها ما يستخدم عناوين التسجيلات الداخلية للعرض المباشر أمام المستفيد، كما أن هناك ما يستخدم النماذج المصورة في حالة استرجاع المعلومات متعددة الوسائط

(الصور، الأفلام)، وتعمل أغلب أدلة البحث على تحويل عناصر التصفح سواء أكانت مصطلحات أو عناوين إلى روابط ينتقل المستفيد بالضغط عليها مباشرة إلى النتائج المسترجعة.

كما استخدمت كل أدلة بحث عينة الدراسة نمط المصطلحات أو رؤوس الموضوعات لتقسيم مجموعاتها الداخلية، غير أن استخدام المصطلحات هنا لم يكن موفقًا، وذلك لأسباب هي؛ أن المصطلحات المستخدمة كانت عامة جدًا ولا تتيح للمستفيد التعرف إلى مضمونها الموضوعي بدقة كافية، كما أن المصطلحات التي استخدمتها هذه الأدلة لم تتعدَّ مستويين أو ثلاثة مستويات في التخصص الموضوعي، وكان أهم الأسباب هو عدم وجود أي دليل بحث بالعينة يعتمد على تقسيم موضوعي منهجي أو منطقي، بل كان الاعتماد على تقسيم عام يدور حول المرأة والصحة والسفر والألعاب والعلوم، بما لا يستطيع المستفيد معه تتبع المفاهيم الدقيقة التي يشملها كل مصطلح.

ثانيًا: أنظمة التقسيم الموضوعي:

لم يتبع أي من أدلة بحث العينة نظامًا معياريًا منطقيًا من أنظمة التقسيم الموضوعي أو أنظمة التصنيف المعيارية للمعرفة البشرية كخطط التصنيف المرتبة منطقيًا أو قوائم رؤوس الموضوعات المرتبة هجائيًا. وكان الاعتماد السائد على تقسيم عام خالٍ من أي منهجية أو منطقية في تقسيم المصطلحات وترتيبها داخل قائمة التصفح، واختلفت أدلة بحث العينة في مستويات بناء هذه التقسيمات العامة؛ فمنها ما توقف عند حد التقسيم العام والانتقال مباشرة منه إلى المواقع التي تغطي موضوعات هذا التقسيم، ومنها ما جعل التقسيم مدخلاً أوليًا لتغيير نمط التصفح من التصفح بالمصطلحات إلى تصفح المقالات وأعداد الدوريات كما حدث في دليل findarticles الذي عمل على تقسيم الدوريات الرقمية داخله إلى أقسام موضوعية عامة، ثم تم ترتيب عناوين الدوريات تحت كل قسم ترتيبًا

هجائيًا ثم رتبت أعداد كل دورية ترتيبًا زمنيًا بحسب تواريخ نشرها. ومن ثم يمكن القول إن التقسيم المتبع في دليل findarticles هو أفضل أنظمة الترتيب العامة المتبعة داخل أدلة عينة الدراسة.

٢/٣/٥ تكشيف قواعد البيانات وتصنيفها:

يوجد مستويان من مستويات تحليل وتكشيف محتوى قاعدة البيانات يمكن اتباعهما عند بناء أدلة بحث قواعد بيانات الويب غير المرئية ؛ أما **المستوى الأول** فيتمثل في التكشيف أو التحليل العام : وهذا يتم باستخدام المصطلحات الموضوعية والكلمات الكشفية للتعبير عن المجال أو التخصص الموضوعي لقاعدة البيانات كلها، وليس مجموعاتها أو مقالاتها الداخلية. ويستخدم هذا المستوى في دليل البحث حتى يتم إدراج عناوين قواعد البيانات تحت المصطلح أو المصطلحات الموضوعية في قوائم التصفح التابعة لدليل البحث، ومن ثم يتمكن المستفيد من الوصول لقواعد البيانات الملائمة إذا ما اختار مصطلحًا أو أكثر من هذه المصطلحات. أما **المستوى الثاني** فهو التكشيف الداخلي لمجموعات قاعدة البيانات : وهنا يتم تكشيف وتحليل محتوى كل مقالة أو وحدة داخلية داخل قاعدة البيانات، والهدف من ذلك هو أن يتمكن المستفيد من الوصول إلى استرجاع المقالات أو الوحدات الداخلية دون الوقوف عند قاعدة البيانات ذاتها، ويصلح هذا المستوى أكثر في حالة اعتماد دليل البحث على أسلوب البحث الحر إلى جانب قوائم التصفح بالمصطلحات أو الكلمات الدالة.

الجدول (٥-١٥) تكشيف قواعد البيانات وتصنيفها

<table>
<tr><th colspan="2" rowspan="2">تكشيف وتصنيف قواعد البيانات</th><th colspan="6">أدلة بحث الويب غير المرئية</th></tr>
<tr><th>Academici-nfo</th><th>beaucoup</th><th>Complete-planet</th><th>Find-articles</th><th>genealogyspot</th><th>Mag-portal</th></tr>
<tr><td colspan="2">الحصر والترتيب</td><td>---</td><td>---</td><td>√</td><td>√</td><td>---</td><td>√</td></tr>
<tr><td rowspan="2">التكشيف</td><td>آليًا</td><td>---</td><td>---</td><td>---</td><td>---</td><td>---</td><td>---</td></tr>
<tr><td>يدويًا</td><td>√</td><td>√</td><td>√</td><td>√</td><td>√</td><td>√</td></tr>
</table>

يتضح من الجدول (٥-١٥) الملاحظات التالية:

أولاً: اعتمدت مختلف أدلة بحث عينة الدراسة على اتباع مستوى التكشيف العام لقواعد البيانات، وتمثل ذلك في أن أغلب النتائج المسترجعة لهذه الأدلة كان متمثلاً في استرجاع مواقع أو قواعد الويب غير المرئية ذاتها، دون القدرة على استرجاع محتوياتها الداخلية. ويدل ذلك على أن عمل أغلب أدلة البحث وقف عند حد الحصر والترتيب دون الوصول إلى حد التكشيف والتحليل للمحتوى غير المرئي لقواعد بيانات ومواقع الويب غير المرئية.

ثانيًا: عملت أدلة البحث magportal, findarticles, completplanet على إخضاع المجموعات الداخلية لقواعد البيانات إلى التكشيف والتحليل الموضوعي، مما نتج عنه قدرة المستفيدين على بحث واسترجاع المقالات والتسجيلات الداخلية لقواعد البيانات. وقد قدم دليلا البحث magportal, findarticles, أفضل نموذج في القدرة على أداء التكشيف والتحليل الموضوعي الداخلي لمقالات ومجموعات قواعد بيانات الدوريات الرقمية المتخصصة على الويب.

٤/٥ خاتمة:

عمل الفصل الخامس على معاملة وملاحظة واقع عمل أدوات بحث الويب غير المرئية الحالية، من محركات وأدلة، وذلك من خلال دراسة أهم جوانب عمل محركات وأدلة بحث الويب غير المرئية في التعامل مع بحث واسترجاع المحتوى غير المرئي. وكانت هذه الجوانب متمثلة في دراسة خصائص وبنية واجهات البحث، إلى جانب ملاحظة واقع عمل نماذج البحث داخلها، كذلك مشاهدة ومراقبة برامج الزاحف وآليات عملها مع قواعد البيانات بحثًا وإضافة، ثم التركيز على أهم الخدمات التي تقدمها أدوات بحث عينة الدراسة.

وقد كشفت هذه المعايشة عن كثير من الجوانب المؤثرة في إمكانات التعامل مع محتوى الويب غير المرئية والقدرة على استرجاع محتويات قواعد بيانات الويب

غير المرئية. وكان أهم ما توصل إليه الباحث من هذه النتائج أنه ما زال هناك قصور في الدلالة والتطبيق المصاحبين لعمل أدوات بحث الويب غير المرئية من محركات وأدلة بحث على السواء. أيضًا فإن الجيل الحالي من أجيال أدوات بحث الويب غير المرئية يمثل الإرهاصات الأولى في خوض الحديث عن استرجاع محتوى قواعد البيانات على الويب وإنارة هذا النفق المظلم. أيضًا لم يوجد على الإطلاق أي من أدلة البحث، يعمل بمنهجية هادفة للتحكم في محتوى قواعد بيانات الويب تنظيمًا واسترجاعًا، وما يوجد الآن على ساحة الويب، إنما هي محاولات جاهدة لكسر حاجز التعامل مع المحتوى العميق الديناميكي للويب غير المرئية.

الفصل السادس

مواصفات محركات بحث قواعد بيانات الويب غير المرئية

٦/٠ تمهيد:

خلصت الدراسة في فصولها السابقة إلى رؤية نظرية وتطبيقية لواقع أدوات بحث الويب غير المرئية، وقد كان من بين الجوانب التي تم دراستها، خصائص عمل قواعد بيانات الويب وبنيتها الداخلية واتجاهاتها في تكوين مجموعاتها الداخلية من ملفات للمعلومات النهائية وتسجيلات كاملة تحوي مختلف عناصر الوصف الفني. ويركز الباحث في هذا الفصل على تشكيل الملامح العامة المقترحة لبناء محركات بحث قواعد البيانات، داعماً في ذلك التعرض لآليات العمل المختلفة داخل محركات البحث، فضلاً عن الصبغة المثلى التي سيكون عليها برنامج الزاحف، بالإضافة إلى جوانب التحليل والتكشيف التي تتعامل بها محركات بحث الويب غير المرئية مع مجموعات التسجيلات الداخلية وملفات النصوص الكاملة. وسوف يستعرض هذا الفصل ملامح عمل نظم استرجاع المحتوى غير المرئي لقواعد البيانات مركزًا على أربعة أركان أساسية :

الأول : يختص بالجوانب الفنية التي سيتم على أساسها معالجة وتكشيف المحتوى الداخلي لقواعد بيانات الويب.

الثاني : سيركز على خصائص عمل برنامج الزاحف خاصة ومحرك البحث عامة، وأهم الآليات والخصائص الفنية التقنية التي يجب أن يتصف بها محرك البحث.

الثالث : يركز على بناء التقسيمات الموضوعية وقوائم التصفح داخل أدلة بحث قواعد بيانات الويب غير المرئية، حيث سيمثل الدليل هنا جزءاً لا يتجزأ من نظام استرجاع قواعد البيانات المقترح.

الرابع : سوف يتعرض إلى أهم القضايا المتعلقة بعمليات تنظيم واسترجاع المحتوى غير المرئي على الويب، ومن خلال معالجتها يحاول الباحث أن يطرح لها بعض الجوانب الإيجابية الداعمة لعملية تنظيم واسترجاع مجموعات قواعد بيانات

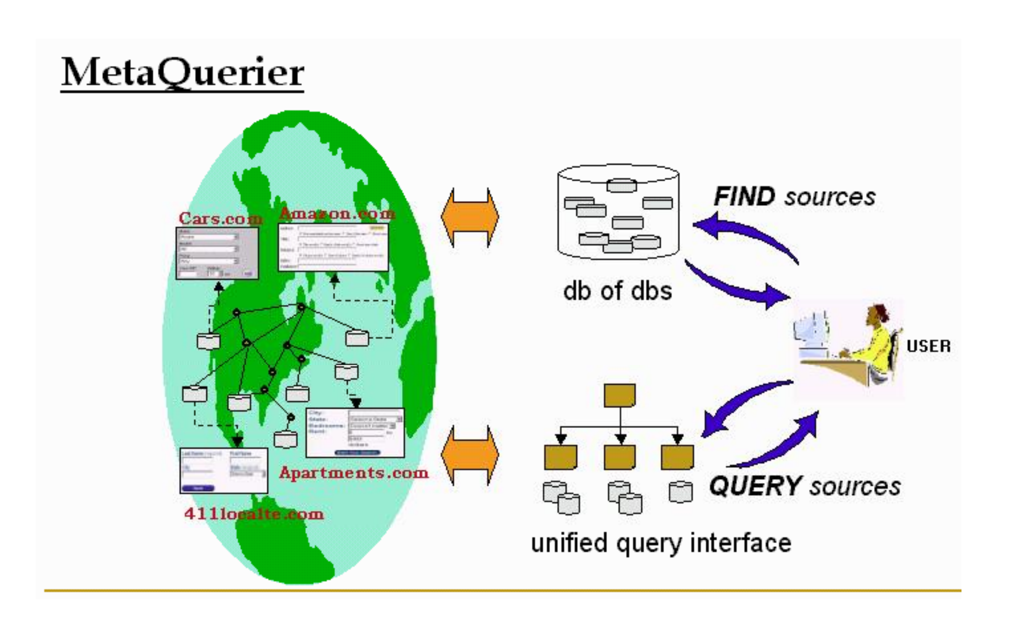
الويب. وأما المبدأ الذي يتبناه الباحث في هذا الفصل فهو أن محرك بحث الويب غير المرئية المقترح سوف يكون أقرب لأداة البحث (بوابات الويب) أكثر منها لمحرك أو دليل بحث. وسوف يطلق الباحث على محرك البحث المقترح مسمى (STORM : the Invisible Web Search Engine) على أن تكون مختلف المواصفات وعناصر البناء الواردة في هذا الفصل هادفة إلى رسم ملامح هذا المحرك المقترح.

٦ / ١ معالجة وضبط قواعد الويب غير المرئية :

تمثل قاعدة بيانات الويب هدفا مزدوج التنظيم والاسترجاع لمحركات بحث الويب غير المرئية ؛ حيث يكون التفاعل في التنظيم والاسترجاع على مستويين هما :

أولاً : تنظيم واسترجاع قاعدة البيانات جميعها.

ثانيًا : تنظيم واسترجاع محتوى وملفات قاعدة البيانات. وفي الحالة الأولى يقف محرك البحث عند استرجاع قاعدة البيانات ذاتها وتنظيمها كوحدة واحدة داخل مجموعات بوابة أو نظام استرجاع الويب غير المرئية، ويختص دليل البحث بهذا الجزء حيث سيعمل على حصر وترتيب قواعد البيانات التي يتعامل معها النظام في شكل قوائم أو تقسيمات موضوعية مرتبة في مستويات متعددة من الموضوعات العامة إلى المتخصصة. أما الحالة الثانية فإن قاعدة البيانات سوف تسمح لمحركات البحث بالتعمق إلى محتوى التسجيلات الببليوجرافية والقدرة على الوصول للملفات المعلومات النهائية، ويرى الباحث أن هذا الجزء سوف يختص به أسلوب البحث الحر بالكلمات المفتاحية من خلال توجيه استفسارات البحث إلى قاعدة البيانات والرد عليها بالنتائج المسترجعة. ويوضح الشكل (٦- ١) طريقتي تعامل المستفيد مع استرجاع قواعد البيانات . ومن ثم فإن الباحث يقترح تبني مستويي العمل السابقين للتنظيم واسترجاع محتوى قواعد البيانات غير المرئي، وبذلك يكون وصف العمل المقترح في العناصر التالية:



الشكل (٦-١) مستويا التنظيم والبحث لقواعد البيانات

١/١/٦ أنماط تكشيف وتحليل محتوى قواعد بيانات الويب:

يستخدم مصطلح تصنيف محتوى قواعد بيانات الويب أكثر من مصطلح التكشيف للتعبير عن التحليل الموضوعي للمجموعات الداخلية لقاعدة بيانات الويب. وذلك يرجع في رأي الباحث إلى أن محركات بحث الويب غير المرئية لا تهدف إلى تكشيف التسجيلات أو الوحدات الداخلية لقاعدة البيانات لكي تضمها إلى قاعدة المحرك، وإنما الهدف من التحليل الموضوعي لمجموعات قاعدة البيانات هو التعرف إلى المجالات الموضوعية التي يمكن أن تدرج أسفلها قاعدة بيانات الويب، ومن ثم يمكن إخضاعها للبحث عنها وتوجيه استفسارات المستفيدين إليها. وعلى ذلك فإن تعامل أدوات بحث الويب مع قواعد البيانات يأتي في سياق التحليل الموضوعي بغرض التصنيف، وليس بغرض التعريف بالمحتوى في مصطلحات وهو ما يعرف بعملية التكشيف.

١/١/١/٦ معالجة قاعدة البيانات واسترجاعها:

اتجهت أدوات بحث عينة الدراسة إلى حصر وترتيب قواعد بيانات الويب التي تتعامل معها، وهي بذلك تعتمد على مبدأ الحصر والإتاحة لمسميات قواعد البيانات أمام المستفيدين ؛ مما يعني أن هذه الأدوات تكتفي بدورها كأدلة أو أدوات حصر وليس بحثًا مباشرًا لمحتوى قاعدة البيانات. ومن ثم فإن هذا المستوى من التحليل العام لمجال وطبيعة قاعدة البيانات يقتضي معه أن تُتيح أدلة البحث بعضَ المعلومات الواصفة لقاعدة بيانات الويب. وهذا الوصف الفني سيكون بمثابة التسجيلات الببليوجرافية الملحقة بكل قاعدة بيانات للتعبير عنها والتعريف بها أمام المستفيد قبل أن يتجه للبحث داخلها. غير أن بناء هذه التسجيلات سوف يعتمد إلى حد كبير على مراجعة وتقييم العنصر البشري لهذا الإجراء ؛ وذلك لأنه لا توجد معايير موحدة تتعامل مع قواعد بيانات الويب على كونها مصدرًا للمعلومات، بل إن التعامل مع قواعد بيانات الويب يسير دائمًا في اتجاه أنها أحد نظم استرجاع المعلومات الرقمية على الويب.

ويمكن أن تتكون تسجيلات قواعد البيانات وعناصر الوصف بها كالتالي(١):

الجدول (٦-١) حقول التسجيلة الببليوجرافية عن قاعدة البيانات داخل الدليل

Database record fields	حقول تسجيلة قاعدة البيانات
Title	عنوان قاعدة البيانات.
Agency/Owner	المؤسسة أو المالك.
Database Abstract	مستخلص عن قاعدة البيانات.
Database Update Frequency	دورية وفترات التحديث لقاعدة البيانات.
Legal Constraints on Database Access	خصائص الإتاحة الشرعية للوصول لقاعدة البيانات.

(١) Google and the deep web. Blogger, 2007, cited 1/7/2008, cited at http//glinden.blogspot.com/2007/03/google-and-deep-web.html

حقول تسجيلة قاعدة البيانات	Database record fields
الأشكال الرقمية لقاعدة البيانات.	Form(s) of Digital Copies of Database
تكاليف النسخ أو مقابل التحميل.	Database Digital Copy Media and Costs List
تكاليف إعادة الإنتاج لقاعدة البيانات.	Report Reproduction Cost List
تقرير مقابل الحصول على خدمات قاعدة البيانات.	Custom Services Costs
قائمة نماذج البيانات داخل قاعدة البيانات.	List of Database Data Modules
قائمة حقول التسجيلات داخل قاعدة البيانات	Annotated List of Fields (one list per data module)

وعلى ذلك فإن الإجراءات المتبعة في معالجة قواعد بيانات الويب جميعها كالتالي:

١- البحث والتعرف إلى قواعد بيانات الويب المختلفة في الموضوع، سواء أكان ذلك باستخدام الزاحف أم عن طريق العنصر البشري.

٢- تحليل المجال الموضوعية لقاعدة البيانات والتعرف إلى خصائص العمل وبنية المعلومات.

٣- صياغة مختلف خصائص وعناصر وصف المجال الموضوعي والتقني في شكل عناصر أو حقول وصف ببليوجرافي الجدول (٦-١).

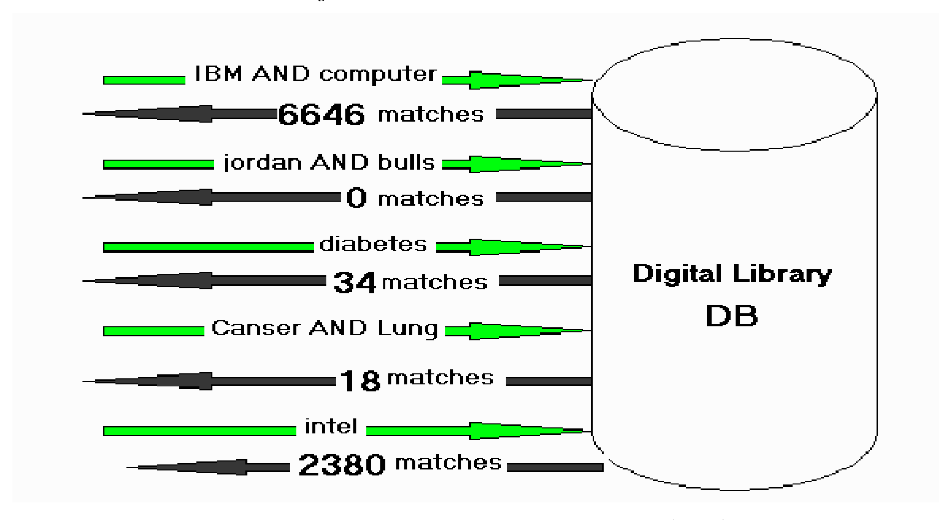
٤- ربط كل قاعدة بيانات على حدة بالمصطلح أو رأس الموضوع المتوافق معها في قوائم رؤوس موضوعات دليل البحث.

٢/١/١/٦ معالجة واسترجاع محتوى قاعدة البيانات:

يمثل محتوى قواعد بيانات الويب القسم الأكبر من معلومات محتوى الويب غير المرئي، وتأتي مواقع الويب التفاعلية ومتغيرة المحتوى في المرتبة الثانية. ومن ثم فإن محركات بحث الويب غير المرئية تهدف إلى طرق باب هذا المحتوى بمجموعة من الاستفسارات البحثية التي يطرحها برنامج محدد حتى يمكن استقراء

الموضوعات المتخصصة لمجموعات قاعدة البيانات وتصنيفها اعتمادًا على تصنيف النتائج المسترجعة.

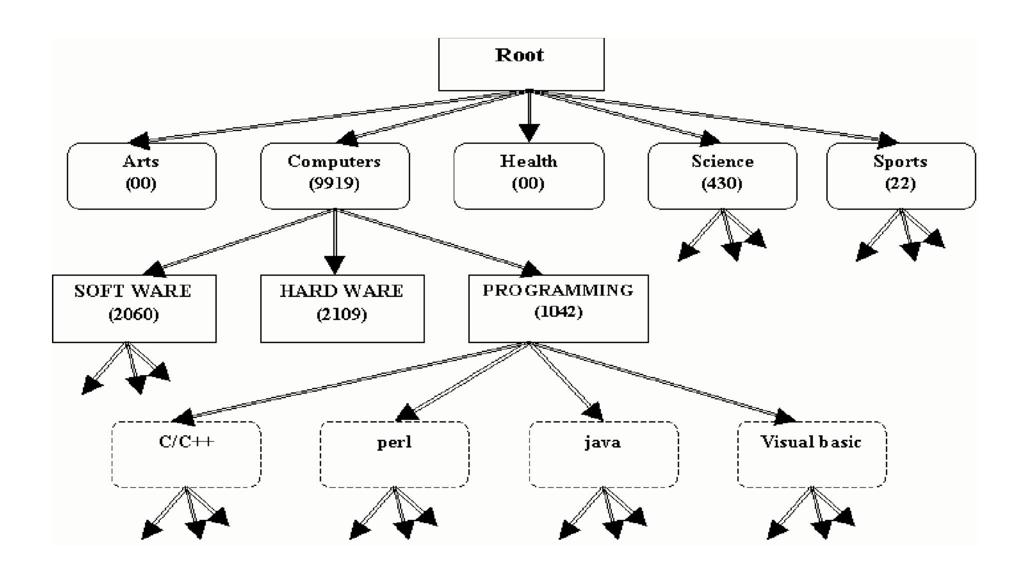
وقد تبنت جامعة كولومبيا Columbia University نموذج Qprober لتحليل محتوى قواعد البيانات النصية، مستخدمة أسلوب طرح الاستفسار وتحليل النتائج المستدعاة. يطرح هذا النظام كلماته البحثية على قاعدة البيانات للتعرف إلى وحدات النتائج المستدعاة، ويستقبل البرنامج الفرعي classifiers هذه النتائج ثم يحللها ويتعرف إلى موضوعاتها، ثم يضعها في تقسيم هرمي للموضوعات التي توجد بداخله. غير أن التحليل لا يكون من خلال مضاهاة الكلمات، وإنما من استقراء أعداد النتائج التي تخرج مع كل مصطلح بحث. وتعكس أعداد النتائج المستدعاة ترتيب قاعدة البيانات في الشكل الهرمي للموضوعات. ويوضح الشكلان (٦-٢) و(٦-٣) استخدام مصطلحات البحث واستقراء الأعداد المسترجعة، ثم مكان قاعدة البيانات في التوزيع الهرمي للموضوعات[١].



الشكل (٦-٢) استخدام مصطلحات البحث واستقراء الأعداد المسترجعة

---

(١) Gravano, Luis and Ipeirotis, Panagiotis G. QProber: A System for Automatic Classification of Hidden-Web Databases, Columbia University, 2003, cited 1/8/2007, cited at htt//qprober.cs.columbia.edu/publications/tois2003.pdf



الشكل (٦-٣) مكان قاعدة البيانات في التوزيع الهرمي للموضوعات

ويرى الباحث أن استخدام هذا النموذج لا يقتصر فقط على قواعد البيانات النصية، وإنما يمكن استخدامه مع قواعد بيانات الويب متعددة الوسائط ؛ من خلال طرح كلمات البحث وتطابقها مع الكلمات المفتاحية في الميتاداتا metadata، بعيدًا عن التعرض لمحتوى الصوت أو الصورة. ومن ثم يمكن استخدام هذا النظام مع قواعد بيانات الويب غير المرئية عامة. بالإضافة إلى أن استخدام هذا النموذج قاصر على تكشيف قواعد البيانات وتصنيفها، وليس لتصنيف أو تكشيف محتواها الداخلي فيما بين مجموعات البيانات بعضها بين بعض. أي أنه يستخدم للتعامل مع قاعدة البيانات كمصدر وليس كنظام استرجاع. وتعتمد أغلب نظم إدارة قواعد البيانات على أسلوب البحث بالكلمات المفتاحية دون اللجوء إلى أسلوب التصفح أو الإبحار لاسترجاع محتوياتها، ولذلك فلا تجد هذه النظم سوى الاعتماد على نظام تكشيف لمحتوياتها من تسجيلات المعلومات. ويعد نظام B+ trees من أكثر النماذج استخدامًا مع قواعد البيانات العلائقية RDBMS مثل IBM, Oracle, Sybase, and Tandem. ويعمل هذا النموذج على إنتاج

شجرة موضوعية للكلمات الدالة داخل مجموعات البيانات، مع ربط كل من التسجيلات بكلمات البحث في مستوياتها الموضوعية المختلفة من خلال row ID (RID) or record identifier[(١)].

٢/١/٦ نماذج بحث واجهات قواعد البيانات:

تتخذ قواعد بيانات الويب من واجهات البحث نوافذ للتعامل مع مستخدمي الويب، ومن خلال هذه الواجهة يستطيع المستفيد فهم خصائص وطبيعة عمل قاعدة البيانات ومن ثم طرح استفسارات البحث وتلقي النتائج المسترجعة. على الجانب الآخر فإن نافذة تعامل قاعدة البيانات مع أدوات بحث الويب غير المرئية، تكمن في نموذج البحث، وهو بمثابة أساس بناء وتصميم واجهة بحث قاعدة البيانات، ومن خلال نموذج البحث يستطيع برنامج الزاحف الوصول إلى قاعدة البيانات ثم تحليل العناصر والآليات المصاغة داخل نموذج البحث. إن التفاعل بين نموذج البحث وبين محرك البحث يتم في مرحلتين هما : الأولى : تأتي أثناء الزيارة الاستكشافية الأولى التي يقوم بها برنامج الزاحف عند الكشف عن هذا النموذج ومحاولة التعرف إليه وتحليل بنيته وخصائصه، أما الثانية فهي عند توجيه الاستفسارات البحثية من محرك بحث قواعد البيانات إلى نموذج البحث حتى يتم بعد ذلك تلقي النتائج المسترجعة وعرضها أمام المستفيد. ولكي يتحقق التفاعل الناجح بين قواعد بيانات الويب وبين محرك البحث المقترح في هذه الدراسة، فسوف يتم الحديث عن أهم الجوانب اللازمة في نموذج بحث واجهة قاعدة البيانات فيما يلي:

١/٢/١/٦ مكونات نموذج البحث الأمثل:

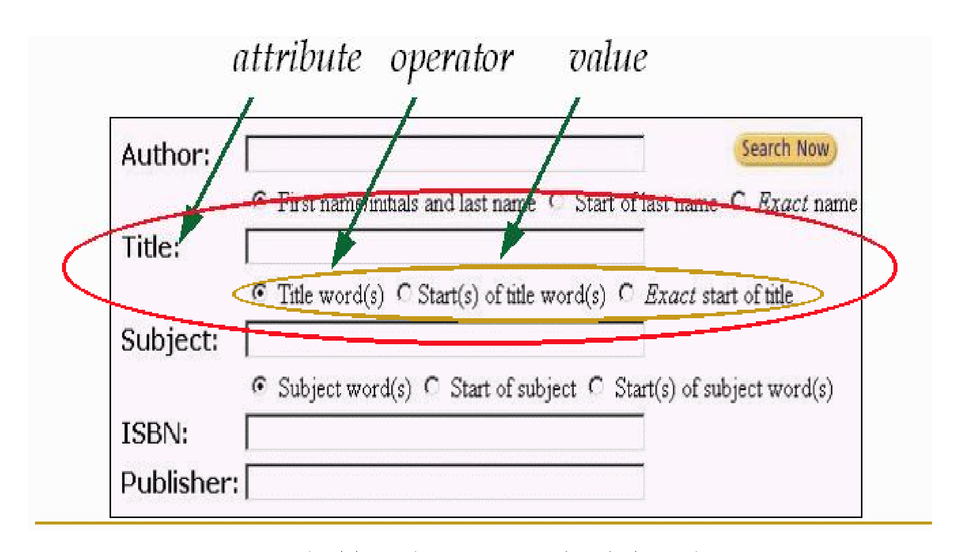
تبدأ برامج الزاحف في تتبع مسارات نماذج البحث محاولة ضم مختلف النماذج التي تصل إليه، ومما لا شك فيه أنها لا تضم كل ما تقابله من نماذج البحث، وإنما تهتم فقط بالنماذج الأكثر إيجابية في إجراء عمليات البحث فيما بعد. ولكي يتم الحكم على النماذج المرجحة وغير المرجحة ؛ فإن برامج الزاحف تعمل

(١) Lauridsen, Helle. Scientist in the Electronic Library-is Deep Indexing of Use for Research?cites , cited at http://www.ffos.hr/lida/datoteke/LIDA2007-Lauridsen_monday.ppt

على تحليل النموذج إلى مكوناته الأساسية حتى يتم تحليل النموذج كله، وهذه المكونات تتمثل في ثلاثة عناصر أساسية هي :

- **عناوين البحث أو صفات البيانات** attributes : وهي المسميات التي تعرف بعناصر البحث مثل المؤلف أو العنوان أو الناشر... إلخ.
- **القيم المدخلة** values : وهي قيم البحث المعلوماتية التي تدخل إلى محتوى قاعدة البيانات.
- **معاملات عناصر البحث** operators : وهي العناصر الفرعية لأحد حقول البحث داخل النموذج.

ويرى الباحث أنه كلما اتجهت نماذج بحث واجهات قواعد بيانات الويب إلى المعيارية والتوحيد في آليات وعناصر البحث، استطاع برنامج الزاحف الوقوف وإجراء المفاضلة والترجيح بين نماذج بحث الويب بأفضل أداء. ويوضح الشكل (٦-٤) صورة مبسطة لنموذج البحث داخل إحدى قواعد بيانات الويب[١].

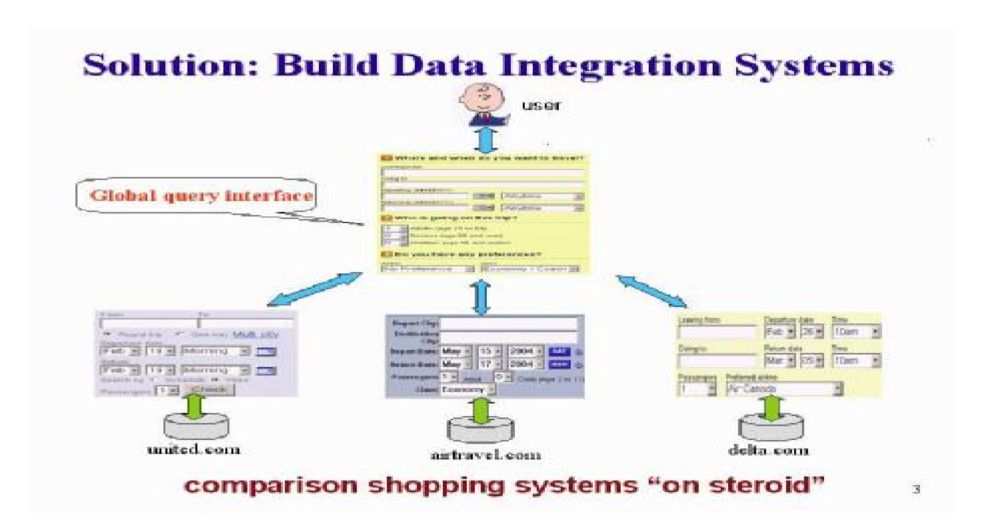


الشكل (٦-٤) نموذج بحث قواعد بيانات الويب

(١) Barbosa, Luciano. Combining Classifiers to Identify Online Databases, University of Utah, 2005, cited at 20/5/2009, cited at http://www2007.org/htmlpapers/paper603/

٢/٢/١/٦ خصائص عمل نموذج بحث قواعد البيانات:

ثمة تطابق بين نموذج بحث قاعدة البيانات وبين النمط الذي صيغت عليه بنية التسجيلات والملفات الداخلية، ومن ثم فإن لكل قاعدة بيانات آليات وعناصر البحث الخاصة بها، التي يعبر عنها في النهاية نموذج البحث، وعلى ذلك فإن برنامج الزاحف سوف يقابل كثيرًا من الأنماط والمسميات المختلفة داخل نماذج بحث قواعد بيانات الويب. ويقوم برنامج الزاحف بإجراء المقابلة والمطابقة بين عناصر البحث في كل نموذج ثم استنباط النموذج العام الذي يعمل به محرك بحث الويب غير المرئية. ويوضح الشكل (٦-٥) عملية المطابقة بين نماذج البحث والوقوف على الشكل الموحد الذي يمكنه التعامل مع أكثر من قاعدة بيانات[١].



الشكل (٦- ٥) المقابلة بين النماذج المختلفة لصياغة النموذج الموحد

وقد ذكرت بعض الدراسات كدراسة Denis Shestakov مدى الاهتمام الذي يتطلبه نموذج البحث عند بنائه داخل صفحات الويب، حيث يمكن للمستفيد

(١) Wu ,Wensheng WebIQ: Learning from the Web to Match Deep-Web Query Interfaces, University of Illinois,2006, cited at 1252009, cited at http://www.cs.binghamton.edu/~meng/pub.d/icde06-webiq.pdf

التعامل معه وملء القيم الملائمة للبحث، ويطلق على هذا النموذج مصطلح html form. وتحوي صفحات الويب هذا النموذج مبتدأً ومنتهيًا بتيجان <form>. وتعمل لغات التحويل هنا على متابعة التعامل مع نموذج البحث ؛ حيث تتلقى هذه اللغات المعلومات الواردة من نموذج البحث ثم تجري عمليات المعالجة لاستفسار البحث، متابعة لاسترجاع النتائج في نماذج أخرى تبنى كواجهات عرض النتائج. وقد أشارت هذه الدراسة أيضًا إلى طبيعة بناء حقول نموذج البحث form fields وهي مثل (checkboxes, radio buttons, menus, etc.) ؛ حيث تحتوي هذه الحقول على ملامح محددة يجب الاهتمام بها عند بناء نموذج البحث جميعه، وتتمثل في[1]:

١- مجال الحقل field domain : وهو كم المعلومات المنتظر إدخاله في هذا الحقل وطبيعته.

٢- مسمى أو عنوان الحقل field label : وهو عنوان الحقل الموضح لهويته عن باقي حقول النموذج.

٣- القيم الذاتية Initial field set : وهي المعلومات الذاتية التي تملأ الحقل إذا لم يدخلها المستفيد.

٣/١/٦ بنية ووصف ملفات "محتوى" قواعد البيانات:

مكنت التقنيات المتقدمة وقدرات العمل والتخزين العالية قواعد بيانات الويب من أن تتعدى مستوى التسجيلات الببليوجرافية، إلى مستوى ملفات المعلومات النهائية التي تحوي النصوص أو الصور أو الصوت. وقد أخذت هذه الملفات طريقها لغزو مواقع الويب عامة، وليس قواعد بيانات الويب فقط. وقد شكلت هذه الملفات جانبًا كبيرًا من مشكلة المحتوى غير المرئي أمام برامج الزاحف ومحركات البحث ؛ حيث تتشكل هذه الملفات في أنواع مختلفة لملفات الويب مثل؛ doc, pdf. Mdb, ppt ... إلخ. ولاحتياج هذه الملفات إلى برامج خاصة لقراءتها

(١) Shestakov, Denis and Bhowmick, Sourav S. DEQUE: querying the deep web, Nanyang Technological University, Singapore, 2004, cited at 2/4/2009, cited at www.sciencedirect.com

وعرضها فإن محركات البحث لم تكن تتعامل مع هذه الملفات من قبل، فضلاً عن أن هذه الملفات لا تمتلك البيانات الوصفية metadata الخاصة بها التي تمكن محرك البحث من تحديد هويتها والتعرف إليها.

يختلف حال ملفات المعلومات النهائية في قواعد البيانات عنه في مواقع الويب الأخرى ؛ لأن هذه الملفات قد لا تجد في مواقع الويب البيانات الواصفة لها والدالة على طبيعتها الموضوعية أو الشكلية، أما في قواعد بيانات الويب فإن ملفات المعلومات النهائية تُعالج فنيًا، بحيث نجد لكل ملف تسجيلة ببليوجرافية تدل عليه، ومن ثم تستطيع أدوات بحث الويب غير المرئية التعرف إلى طبيعة كل ملف وتحديد المحتوى الموضوعي له من خلال عناصر الوصف المرتبطة به. وقد تمكنت بعض أدوات بحث الويب المرئية مثل محرك البحث Google من تحقيق التفاعل مع هذه الملفات إضافةً وتنظيمًا واسترجاعًا، وذلك عن طريق تخصيص البحث في برنامج الزاحف والتعامل مع نهايات مسارات الويب URLs. ولما يملكه ملف المعلومات داخل قاعدة البيانات من أهمية، فإن محركات بحث قواعد البيانات يجب أن تحقق الوصول والاستدعاء لهذه الملفات. ولذلك فإن الباحث يقترح أن تعمل قواعد بيانات الويب على إبراز أماكن تواجد هذه الملفات داخل تسجيلاتها الببليوجرافية؛ بحيث ينصب محتوى حقول الوصف على التعريف بهذا الملف والبيانات الببليوجرافية الخاصة به، أما من الناحية التقنية فإن الباحث يرى ضرورة الاهتمام بتقنيات لغات التحويل أو لغات العرض التي تحقق التوافق بين نظام إدارة قاعدة البيانات وبين متصفح الإنترنت الذي يعمل من خلاله المستفيد، وأهم هذه اللغات CGI, PHP, Active Server Pages (ASP) , VBScript and ActiveX, Java / JDBC / JSQL.

إن التفاعل بين محركات بحث الويب غير المرئية وبين قواعد البيانات يمر بمرحلتين أساسيتين، ففي مرحلة البحث والاسترجاع قد لا يتأثر التفاعل كثيرًا بإهمال أو اهتمام قواعد بيانات الويب بملفاتها ؛ لأن محركات البحث هنا لن

تتكلف سوى إرسال استفسارات البحث إلى قاعدة أو قواعد بيانات الويب المتوافقة وموضوع البحث، ثم تتلقى هذه المحركات مجموعات النتائج المستدعاة من مختلف قواعد البيانات، ومن ثم فإن محركات البحث لن تقترب من ملفات المعلومات النهائية. لكن التأثير الفعلي سوف يأتي في مرحلة التكشيف والتحليل لمحتوى ومجموعات قواعد بيانات الويب عندما تكون هذه الملفات ضمن التسجيلات الداخلية التي تخضع للتحليل والتصنيف والتكشيف من جانب محرك البحث.

٤/١/٦ بنية التسجيلة الموجهة إلى محرك بحث الويب غير المرئية:

تتدرج مكونات قاعدة البيانات من الأكبر ابتداء بالملفات والجداول وصولاً إلى التسجيلة ثم الحقول كأصغر وحدة في بناء قاعدة البيانات. ويهتم هذا الجانب من الحديث عن قواعد بيانات المعلومات بالشكل الذي ستكون عليه التسجيلة الواردة من قاعدة البيانات والموجهة إلى محرك البحث كنتائج مسترجعة تخدم استفسار البحث. ويختلف الأمر في عرض تسجيلات قاعدة البيانات بين العرض على واجهات قاعدة بيانات الويب ذاتها وبين توجيه مجموعات التسجيلات إلى محرك البحث ؛ فقد تميل قواعد بيانات الويب إلى عرض تفاصيل أكثر من تلك المرسلة في كل تسجيلة إلى محرك البحث. وتأتي هذه الإمكانية من أن قواعد البيانات تملك مستويات كثيرة من عرض التسجيلات الببليوجرافية تتمثل في ؛ المستوى الفني وفيه يتم عرض مختلف الحقول التي تتكون منها التسجيلة بما في ذلك الحقول الخاصة بضبط ومسؤولية إدخال بياناتها. أما المستوى الثاني فهو المستوى الكامل وفيه تعرض كل الحقول الخاصة بالوصف الفني الببليوجرافي لملف المعلومات النهائية، والمستوى الثالث هو المستوى المختصر الذي تقتصر فيه التسجيلة على أهم حقول وصف ملف أو مادة المعلومات النهائية.

كما تمتلك قواعد بيانات الويب دون غيرها من قواعد البيانات الأخرى ميزة إضافية بجانب إمكانية العرض في مستويات متعددة. وهي قدرة قواعد بيانات

الويب على إنتاج تسجيلات متعددة باستخدام لغات التحويل تستطيع قاعدة البيانات من خلال هذه التسجيلات المتعددة عرض كل مجموعة متجانسة من حقول التسجيلة الببليوجرافية في نافذة أو إطار عرض مختلف عن باقي الحقول، وهو ما قد يروق لمستخدمي قواعد بيانات الوسائط المتعددة. ويرى الباحث أن التسجيلات المتعددة قد تؤثر سلبًا على تلقي محرك البحث لمجموعات النتائج المسترجعة من قواعد البيانات، ويتمثل هذا التأثر في التعامل مع أنماط مختلفة قد تحتاج إلى إعادة صياغة مرة أخرى، فضلاً عن متطلبات هذه التسجيلات للعرض في واجهات ذات أطر وخصائص محددة قد لا تتواجد في محرك البحث.

ويرى الباحث أن محرك بحث الويب غير المرئية المقترح في هذه الدراسة (STORM) يحتاج إلى تسجيلات مسترجعة ذات مجموعة محددة من الحقول يمكن أن تقتصر على الجوانب التالية :

- ❖ حقل العنوان أو التعريف بوحدة أو ملف المعلومات النهائية.
- ❖ الرابط أو مسار الوصول إلى وحدة وملف المعلومات file URL.
- ❖ العرض المصغر التوضيحي لمادة المعلومات (وهو مخصص للمواد المصورة).
- ❖ حقول أو خيارات التعامل مع التسجيلة، سواء أكانت خيارات تجارية أو خدمية.

٥/١/٦ ترتيب النتائج المسترجعة:

يمثل ترتيب النتائج المسترجعة أهمية خاصة لدى المستفيدين؛ حيث يؤثر نمط الترتيب للنتائج المسترجعة على سهولة تصفحها، فضلاً عن توفير الوقت والجهد للمستفيد في الوصول السريع لأهم المجموعات أو الوحدات التي ترتبط ارتباطًا وثيقًا بموضوع واستفسار البحث. وتزيد درجة أهمية ترتيب النتائج داخل محركات بحث الويب غير المرئية عن غيرها من محركات البحث الأخرى ؛ هذا

لأنها لا تتعامل مع نظام استرجاع أو قاعدة بيانات واحدة، وإنما تتعامل مع قواعد بيانات مختلفة في التنظيم والعرض لمجموعاتها الداخلية. ومن ثم فإن محرك بحث الويب غير المرئية يتأثر بنمط ترتيب النتائج المرسلة إليه من قاعدة البيانات ردًا على استفسار البحث.

لقد أوضحت الدراسة التجريبية لقواعد بيانات الويب في الفصل الرابع أنها تنتهج أنماطًا مختلفة من ترتيب النتائج المسترجعة داخلها ؛ حيث تتنوع هذه الأنماط بين الترتيب بحسب الصلة بموضوع البحث، أو الترتيب بحسب حداثة المعلومات، أو الترتيب بحسب بعض الحقول، مثل المؤلف أو العنوان أو مصدر الاسترجاع، وهناك من قواعد البيانات ما ينتهج أنماطًا أخرى تتشكل بحسب طبيعتها التجارية كالعرض بحسب سعر الوحدة الذي تتبناه قاعدة البيانات Amazon. ومما لا شك فيه أن نوع الترتيب يؤثر على أداء محرك بحث الويب غير المرئية أثناء تلقيه للنتائج المسترجعة من قواعد البيانات ؛ حيث قد يكتفي محرك البحث بعدد محدد من النتائج الموجهة إليه واستبعاد المجموعات الأخرى، أو أن يتم التداخل بين النتائج المسترجعة من قاعدة بيانات. وهذا بدوره يرتبط بطبيعة أعداد النتائج المسترجعة من قواعد البيانات إلى محرك بحث الويب غير المرئية ؛ فكلما زادت أعداد النتائج المسترجعة، أثر ذلك على طبيعة الدمج والفرز والترتيب الذي تقوم به محركات البحث، وهو ما سيتم الحديث عنه تفصيلاً في عنصر لاحق من هذا الفصل أثناء الحديث عن عملية الدمج والفرز داخل محرك بحث الويب غير المرئية.

٢/٦ ملامح بناء محركات قواعد الويب غير المرئية :

يهدف هذا الجزء من الدراسة إلى بناء الأسس الدلالية والاصطلاحية والتطبيقية لمحركات بحث الويب غير المرئية ؛ حيث ما زالت محركات بحث الويب غير المرئية تعاني من عدم دقة المفاهيم والمصطلحات المستخدمة، فضلاً عما يعكسه واقع هذه المحركات من نقص كفاءة العمل والافتقار إلى

إمكانات التنظيم والاسترجاع الملائمة للتعامل مع محتوى الويب غير المرئي. ويعبر الشكل (٦-٦) عن البنية العامة لمحرك بحث الويب غير المرئية. وسوف تطرح الدراسة المفهوم الدلالي والتطبيقي لمحركات بحث الويب غير المرئية في ثلاثة عناصر هي :

١/٢/٦ البناء الدلالي لمحرك بحث الويب غير المرئية:

لم يكن لهذه الدراسة أن تنتهي حتى تضع الأساس النظري للجيل الجديد من محركات بحث الويب. حيث اهتمت أدبيات الموضوع الأجنبية بالتركيز على الجانب الإجرائي في التعريف بهذا الوافد الجديد من محركات البحث، دون محاولة تأصيل مفهوم ثابت يعبر بشكل قاطع عن ماهية محركات بحث الويب غير المرئية. وقد انصب اهتمام مختلف الدراسات على عرض جوانب ومشكلة الاسترجاع لمحتوى ومجموعات الويب غير المرئية. ومن ثم فإن الباحث قد عمل على تبني مفهوم ثابت لمحرك بحث الويب غير المرئية، يجمع في خصائصه التعريف بطبيعة العمل والبنية لهذا النوع من المحركات. ويتمثل هذا المفهوم في النص التالي:

(محركات بحث الويب غير المرئية هي : جيل جديد من أدوات البحث يهتم بتنظيم واسترجاع محتوى قواعد البيانات ومواقع الويب الديناميكية، معتمدًا في ذلك على بناء متطور لبرامج الزاحف المختصة بكشف وتحليل نماذج بحث الويب web search forms. وهي تتطابق في فكرتها وأدائها مع محركات البحث المتعددة).

ويحمل المفهوم السابق بعض الخصائص الأساسية مثل:

١- عدم امتلاك قاعدة بيانات خاصة، إلا في حالة التعامل مع مواقع الويب الديناميكية.

٢- العمل باستخدام برامج ذات بنية مختلف من برامج الزاحف أو العنكبوت.

٣- القدرة على التعامل مع نظم استرجاع مختلفة البنية في آن واحد، كقواعد بيانات الويب.

٤- القيام بدور الوساطة في عمليات البحث بين المستفيد باستخدام واجهة بحث عامة، وبين قواعد البيانات.

٥- التطابق في فكرة العمل بينه وبين محركات البحث المتعددة، غير أنه يعمل باستخدام الزاحف.

٢/٢/٦ النظرة المتقدمة لمحرك بحث الويب غير المرئية:

يقصد الباحث بالنظرة المتقدمة لهذا الجيل من محركات البحث، أن يخرج هذا النوع من محركات البحث عن الإطار الضيق الذي تنحصر فيه كل محركات بحث الويب ؛ والخروج إلى الإطار الرحب الذي لا يعمل فيه محرك البحث كأداة بحث بأسلوب البحث الحر، وإنما يجب أن يكون محرك بحث الويب غير المرئية أداة متكاملة للبحث تجمع بين أسلوبي البحث الحر بالكلمات الدالة، وأسلوب البحث بالتصفح. واستكمالاً للشكل المتكامل لمحركات بحث الويب غير المرئية، فإن ثمة خصائص عمل يجب أن تضاف إلى هذه المحركات مثل؛ تقديم هذه المحركات لبعض الخدمات كالبريد وإنشاء الحساب الخاص بكل مستفيد، بالإضافة إلى الخدمات الداعمة لبحث الويب مثل عرض أهم قواعد البيانات، والتوجه إلى التعامل مع الوسائط المتعددة.

ويمكن تلخيص هذه النظرة المتقدمة لمحركات بحث الويب غير المرئية في أنها يجب أن تعمل كبوابة لبحث الويب غير المرئية invisible web search portal، تجمع في ثنايا عملها كثيرًا من الإجراءات والكثير من الخدمات وأساليب البحث، بحيث تكون هذه البوابة أكبر وأشمل مفهومًا وأداءً من مفهوم وأداء محركات البحث.

٣/٢/٦ مكونات محركات بحث الويب غير المرئية:

لقد تمت الإشارة إلى أن محركات بحث الويب غير المرئية تنتهج منهج محركات البحث في إجراء عمليات البحث والاسترجاع من أكثر من نظام استرجاع مختلف في وقت واحد، كما تمت الإشارة إلى أن محركات بحث الويب

غير المرئية لابد أن تعمل كبوابة للدخول إلى الويب غير المرئية. ولهذا التداخل في المفهوم والأداء فإن من الواجب هنا أن نعرض لمكونات وبرامج العمل المقترحة داخل محركات بحث الويب غير المرئية، وفقاً لمجموعة من قطاعات العمل تتمثل في :

أولاً: قطاع البرامج : وهي مجموعة البرامج التي ستقوم بمهام محددة داخل محرك البحث مثل:

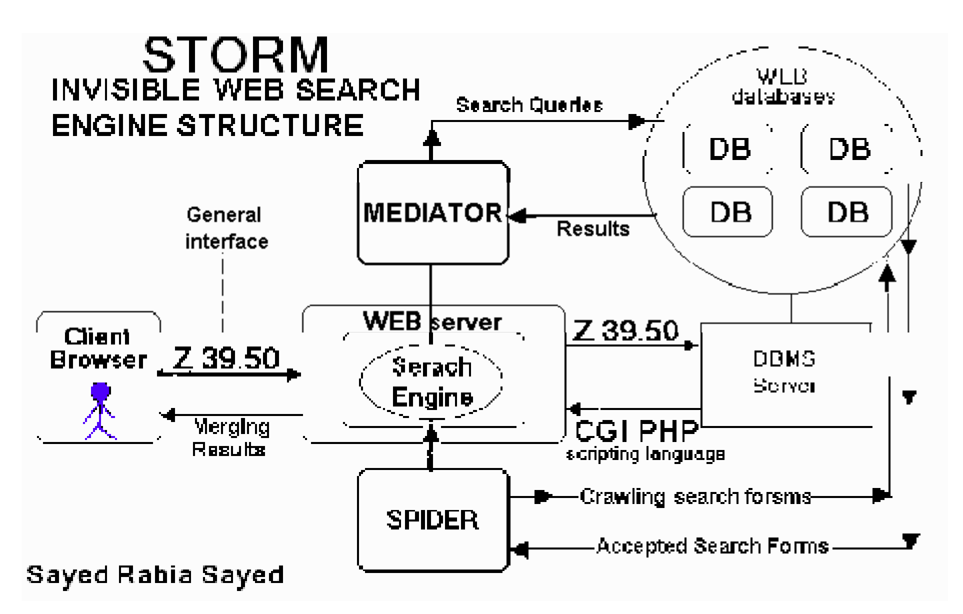
- ❖ برنامج الزاحف لكشف وتحليل نماذج بحث الويب.
- ❖ برنامج الوسيط لإدارة عمليات البحث كلية.
- ❖ آليات وعناصر البحث لتلقي الاستفسارات باسخدام معيار "Z39.50".
- ❖ برامج الفرز والترتيب وتقدير الأهمية.

ثانيًا: قطاع البحث : وهو مجموعة الأساليب العامة المستخدمة لبحث قواعد البيانات، ويضم :

- ❖ أسلوب البحث الحر بالكلمات المفتاحية.
- ❖ أسلوب البحث بالتصفح والتقسيمات الموضوعية.
- ❖ واجهة البحث العامة ونموذج البحث العام، الذي يحتوي على عناصر وحقول بحث محددة.

ثالثًا: قطاع الخدمات : ويضم مجموعة الخدمات المنوط بتقديمها محرك البحث، مثل :

- ❖ الخدمات الداعمة للبحث : مثل البحث داخل النتائج، وإعادة الترتيب، والتحكم بعرض النتائج.
- ❖ الخدمات المكملة للبحث : مثل خدمات بناء ملامح البحث، خدمات حفظ النتائج، والطباعة ... إلخ.



الشكل (٦-٦) البنية العامة المقترحة لمحرك بحث الويب غير المرئية STORM

يوضح الشكل (٦-٦) البنية العامة المقترحة لمحرك بحث الويب غير المرئية Storm، وتبدأ من نافذة متصفح الإنترنت لدى المستفيد ثم محرك البحث، ويمثل الطرف الأول، أما الطرف الثاني فيتمثل في مجموعة قواعد بيانات الويب، ويمثل معيار البحث الموحد Z39.50 وسيلة البحث في نظم استرجاع مختلفة في آن واحد، كما تعمل لغات التحويل على دمج النتائج المسترجعة في شكل واحد وتحويلها إلى نافذة متصفح الإنترنت أمام المستفيد.

١/٢/٦ بنية برنامج زاحف الويب غير المرئية:

١/١/٢/٦ اجتياز الزاحف لنماذج البحث:

يعمل برنامج الزاحف داخل محركات بحث الويب المرئية وغير المرئية على تكشيف وإضافة وتحليل صفحات الويب. غير أن الفارق في العمل بين زاحف الويب المرئية وغير المرئية، يكمن في وجهة العمل في كل من الاثنين ؛ فزاحف

الويب المرئية يعمل باستخدام الروابط الفائقة والمسارات URLs للوصول إلى صفحات الويب ذات بنية html ، والعمل على تحليل الكلمات الدالة الواردة بها في حقول الوصف المختلفة metadata. أما زاحف الويب غير المرئية هنا فيسعى للوصول إلى نماذج بحث مواقع والويب وقواعد البيانات web search forms مستخدمًا لذلك خصائص التعرف إلى حقول البحث وتحليلها ، والتعرف إلى بنية كل نموذج من نماذج البحث المختارة search form fields structure. ويعرف الزاحف عامة على أنه " البرنامج الذي يجتاز الويب بشكل آلي العمل automatically لإضافة صفحات الويب إلى محرك البحث ، ويستخدم في ذلك الروابط الفائقة للتنقل بين صفحات الويب ، غير أنه يفتقر إلى عنصر الروابط الفائقة في التعرف إلى نماذج البحث مع الويب غير المرئية ". وعندما اختلفت بيئة عمل برنامج الزاحف وتحولت من وجهة مرئية ذات روابط فائقة ، إلى وجهة خفية ذات نماذج بحث مستترة ، فقد واجه برنامج الزاحف كثيرًا من الصعوبات في القيام بالمهام المنوطة به داخل محرك البحث ، مثل(١) :

١- القدرة على تحليل ومعالجة نماذج بحث الويب.

٢- القدرة على تكوين وطرح استفسارات البحث على نماذج البحث.

٣- القدرة على التحليل الذاتي للنتائج المسترجعة Automatic response analysis.

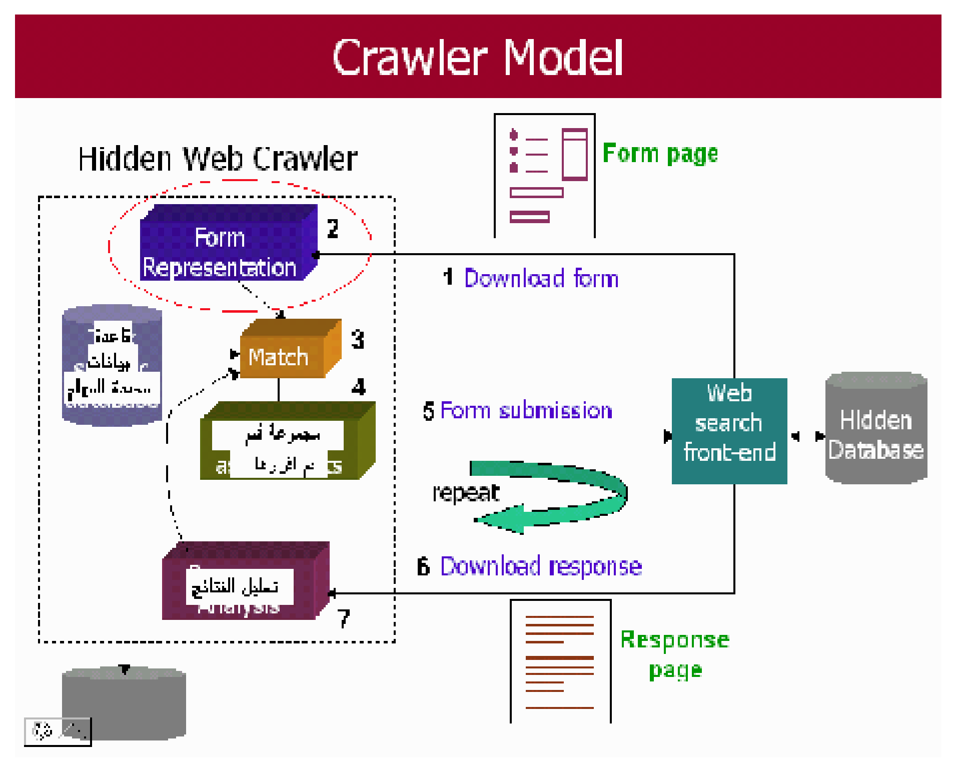
٢/١/٢/٦ مراحل عمل الزاحف المقترحة للويب غير المرئية:

يتناول هذا الجزء من الدراسة التعريف بمراحل العمل المقترحة لبرنامج الزاحف داخل الويب غير المرئية ، وهي المراحل التي تبدأ بالوصول إلى نموذج البحث والتحليل لخصائصه ، انتهاءً بعملية الإضافة إلى محرك بحث الويب غير المرئية.

---

(١) Ntoulas, Alexandros. Crawling and Searching the HiddenWeb, University of California, 2006, cited at 20/5/2009, cited at http://portal.acm.org/citation.cfm?id=1061951.1061952

يمثل الشكل (٦-٧) شكل ومراحل العمل المقترحة لبرنامج زاحف الويب غير المرئية، وهي كما يلي[١]:



الشكل (٦-٧) شكل ومراحل عمل برنامج الزاحف للويب غير المرئية[٢]

(١) El-Desouky I., Ali, Ali, Hesham A. and El-Ghamrawy, Sally M. A new framework for domain-specific hidden web crawling based on data extraction technieques, Mansoura University, Egypt, 2007, cited at 12/4/2009, cited at http://www.icict.gov.eg/ICICT-2006/Papers/Design%20for%20Information%20Retrieval%20Systems%20of%20User%20Interfaces/blind%20final%20ICICT%20A%20New%20framework%20for%20domaiN-specific%20hw%20crawling.pdf

(٢) Raghavan, Sriram. Crawling the Hidden Web & Other Web-related Research,stanford university, 2005, cited at 1/5/2009, cited at http://www.acm.org/fcrc/PlenaryTalks/Garcia-Molina.pdf

٦/٢/١/٢/١ الوصول إلى نموذج البحث:

يصادف برنامج الزاحف أثناء اجتيازه لصفحات ومواقع الويب كثيرًا من الصفحات التي قد تحتوي أو لا تحتوي على نماذج بحث، ومن ثم فإن الهدف في هذه المرحلة هو الوقوف على أهم صفحات الويب التي تحوي نماذج البحث، والتخلي عن تلك التي لا تحمل نماذج بحث. وتقوم آلية محلل النماذج Form Analyzer بالتعرف إلى طبيعة نموذج البحث، وتبيان ما إذا كان يحتوي على نموذج غني بالمعلومات أم لا. وتتباين أنواع نماذج البحث التي يعثر عليها برنامج الزاحف في هذه المرحلة بين نماذج الاشتراك register، ونماذج الشراء purchase، ونماذج الدخول login forms ... إلخ. ولذلك كان هدف هذه المرحلة أن يصل الزاحف إلى نموذج بحث واسترجاع المعلومات لقواعد البيانات ومواقع الويب غير المرئية.

٦/٢/١/٢/٢ تحليل عناصر نموذج البحث:

تنتهي المرحلة الأولى عند إقرار نموذج البحث الذي سيخضع للتحليل والدراسة من جانب برنامج الزاحف، لتبدأ المرحلة الثانية بتحليل تام لعناصر وبنية نموذج البحث Form Analyzing التي تتم على خطوتين هما:

١- محلل النموذج Form parser :

يعمل محلل النموذج على تحديد مجموعة العناصر التي توجد داخل نموذج البحث، وهل هذا النموذج أحادي عناصر البحث أم متعدد عناصر البحث-single attribute or the multi-attribute forms. وترجع الأهمية في ذلك إلى العمل على بنية زاحف متكامل لا يقف فقط عند نماذج البحث الأحادية أو المتعددة، وإنما يملك القدرة على التعامل مع أشكال نماذج البحث كافة. وتتواجد نماذج البحث الأحادية داخل واجهات مواقع الويب التي لا تملك بنية هيكلية أو تنظيم داخل لمقالاتها يعتمد على حقول وصف متعددة، ومن ثم فهي تكتفي فقط ببناء نموذج بحث بسيط يوفر فقط عنصر فراغ البحث text box حتى يتلقى الكلمات

المفتاحية التي سيبحث بها المستفيد. ويختلف هذا النموذج عن نموذج البحث المتعدد الذي غالبًا ما يتواجد في قواعد البيانات التي تعتمد في بناء وتنظيم تسجيلاتها على حقول الوصف المتنوعة، ومن ثم فإنها توفر عناصر بحث عدة داخل نموذج البحث.

٢- اقتباس العناوين والمسميات Label extractor :

يعمل برنامج مقتبس العناوين على تكشيف وتحليل عناوين عناصر بحث النموذج، غير أنه من المهم هنا التأكيد على تيجان اللغة المعيارية html tags ؛ حيث تحافظ هذه التيجان على العناوين وعناصر البحث التي تمثلها، بما لا يدع فرصة للخطأ في تكشيف حقول نموذج البحث. ويوجد نوعان من برامج اقتباس عناوين حقول البحث ؛ أحدهما يختص بتحليل نماذج البحث أحادية العنصر أو الحقل، والآخر يختص بنماذج البحث متعددة الحقول. ويقوم كل برنامج منهما بالعمل بشكل منفصل عن الآخر، بما يجعل نماذج البحث الأحادية في جانب مختلف عن النماذج المتعددة. بالإضافة إلى أنه يتم اقتباس عناوين حقول البحث وفقًا لأشكال مختلفة ؛ فقد يوجد العنوان في يمين حقل البحث، وقد يوجد أيضًا في يساره أو أعلاه، ومن ثم فإن لكل شكل من هذه الأشكال نمطًا مختلفًا في التحليل والاقتباس عن الآخر. ويمكن القول إن عملية تحليل واقتباس عناوين البحث هي المرحلة الفاصلة في عمل برنامج الزاحف، فإذا صلحت، صلحت باقي المراحل، وإذا فشلت، أخفق الزاحف في التعامل مع نماذج بحث قواعد بيانات الويب.

٣/٢/١/٢/٦ برنامج مضاهاة عناوين نموذج البحث Label Matcher:

يعمل برنامج الزاحف على مضاهاة مجموعة العناوين المقتبسة من نموذج البحث متعدد الحقول، مع مجموعات المصطلحات المختزنة داخل قاعدته ذات القيم المتعددة، والهدف من ذلك هو التعرف إلى مجموعة القيم التي تلائم البحث

في هذا النموذج، حتى يتم استخدامها في ملء نموذج البحث وتوجيهه إلى موقع أو قاعدة بيانات الويب غير المرئية. ويعمل برنامج الزاحف في هذه المرحلة على مبدأ مؤداه أن نموذج البحث لابد أن يحمل حتمًا عناوين حقول البحث التي تعبر عن المحتوى والمجموعات الداخلية لقاعدة البيانات التي ينتمي لها، حتى يساهم في الاسترجاع الدقيق لمجموعات هذه القاعدة. ومن ثم فإن مضاهاة عناوين الحقول داخل نموذج البحث هي أفضل وسيلة للتعرف إلى القيم التي يعمل بها نموذج البحث، ويسترجع المطابق لها من قاعدة البيانات.

٤/٢/١/٢/٦ إقرار مجموعة القيم وانتخابها : Form processing

يخلص الزاحف من مرحلة المضاهاة السابقة إلى مجموعة القيم والمصطلحات الملائمة لملء نموذج البحث، وتبعًا لواقع قواعد بيانات ومواقع الويب غير المرئية فإن هذه القيم تختلف من نموذج إلى آخر. فضلاً عن أن برنامج الزاحف يعمل على إعادة وتكرار عملية البحث عن قيم جديدة في كل مرة يرجع فيها إلى نموذج بحث قاعدة البيانات. ويبقى من هذه المرحلة أن برنامج الزاحف إذا لم يجد مجموعة من القيم التي تجمع بين عناوين حقول البحث في النموذج، وبين المصطلحات المتواجدة في قاعدة بياناته، فإنه يهمل أو يلغي الأخذ بهذا النموذج، ساعيًا للعمل مع نماذج بحث أخرى.

٥/٢/١/٢/٦ توجيه استفسار البحث:

يقوم برنامج الزاحف في هذه المرحلة بدور المستفيد ؛ حيث يرسل مجموعة من قيم البحث والكلمات المفتاحية إلى قاعدة البيانات منتظرًا استرجاع مجموعات النتائج المطابقة لاستفسار البحث.

٦/٢/١/٢/٦ استقبال النتائج المسترجعة:

يعمل الزاحف على تحليل واجهات عرض النتائج لتحديد خصائص وطبيعة النتائج المسترجعة ومدى توافقها مع مجموعات القيم المرسلة في نموذج البحث.

٧/٢/١/٢/٦ تحليل النتائج وإضافة نموذج البحث response analyzer :

تمثل عملية تحليل النتائج المرحلة الأخيرة من تكشيف وتحليل نموذج بحث قواعد البيانات، والدور المهم لهذه المرحلة هو التمييز بين مجموعة الصفحات التي تحتوي على نتائج أو وحدات مسترجعة، وبين الصفحات التي تحتوي على تقارير أو رسائل الخطأ التي تفيد بأنه لا توجد نتائج مطابقة للقيم المرسلة في نموذج البحث. ومن ثم يأخذ برنامج الزاحف القرار النهائي بإهمال نموذج البحث أو الإبقاء عليه وتحويله إلى قاعدة بيانات محرك بحث الويب غير المرئية.

ويفيد في هذا الإطار ما حدده محرك البحث google من عناصر يمكن لبرنامج الزاحف اقتباسها عن نموذج البحث أثناء تكشيفه وتحليل خصائصه. وهذه المعلومات تتمثل في(١):

❖ **معلومات عن حقول نموذج البحث**، مثل:

١) عناوين ومسميات الحقول.

٢) نوع حقل البحث، مثل: (فراغ النصوص، القوائم المنسدلة، أزرار الاختيار radio button).

٣) نماذج للقيم المدخلة في حقول البحث (مثل بعض القيم الموجودة في إحدى القوائم المنسدلة).

❖ **فئة نموذج البحث:** حيث توجد بعض نماذج البحث المتعددة التي تبحث في أكثر من جانب في قاعدة البيانات، كأن تخصص إحدى الدوريات على الويب نموذجًا لبحث العقارات، وآخر لبحث الوظائف. أو أن يكون النموذج موحدًا لكامل قاعدة البيانات أو موقع الويب.

(١) Invented by Alon Y. Halevy, Jayant Madhavan, and David H. Ko US Patent Application 20060230033 Published October 12, 2006 Filed on April 5, 2006
http://appft1.uspto.gov/netacgi/nph-Parser?

- ❖ **المعلومات الجغرافية :** كأن يضع أحد الفنادق معلومات جغرافية داخل نموذج البحث عن المنطقة المحيطة، ويمكن عملاءه من استرجاعها.
- ❖ **الحقول الإلزامية** Required fields : وهي مجموعة الحقول التي لا يمكن تفعيل نموذج البحث بدون استكمال قيم البحث داخلها.
- ❖ **خصائص النتائج المسترجعة :** حيث يمكن التحكم في طبيعة النتائج المستدعاة أثناء ملء نموذج البحث.
- ❖ **تخطيط نموذج البحث** Schema mapping : حيث يمكن التعرف إلى الروابط بين حقول نموذج البحث، فضلاً عن العلاقات بين الحقول الرئيسية والفرعية.

٢/٢/٦ إدارة عملية بحث قواعد البيانات داخل محرك البحث STORM:

يعتبر التفاعل بين محرك البحث وقاعدة البيانات مجموعة من الإجراءات المتلاحقة المتكررة دائمًا؛ حيث يبدأ التفاعل بين محرك البحث وقاعدة البيانات عندما يصل برنامج الزاحف إلى نموذج بحث قاعدة البيانات حاملاً رسالة الضم والإضافة إلى مجموعات محرك البحث من قواعد البيانات المختلفة، وعندما ينتهي التعارف بين الزاحف ونموذج البحث بتحليل وتكشيف عناصر وآليات البحث مكونة لهذا النموذج، تبدأ المرحلة التالية من العمل بإجراء عمليات التحليل والتكشيف لمجموعات وتسجيلات المحتوى الداخلي لقاعدة البيانات حتى يتم التعرف إلى ملفاتها وموضوعاتها كي تنضم إلى قوائم الموضوعات في الدليل، وبعد الانتهاء من هذه المراحل تبدأ المرحلة الدائمة المتكررة المتمثلة في توجيه استفسارات بحث المستفيدين إلى قاعدة البيانات، لتلقي النتائج المسترجعة وعرضها أمام المستفيد.

يعتمد محرك بحث الويب المرئية أو غير المرئية على مجموعة من البرامج المتكاملة التي تعمل ذاتيًا لأداء مهام محرك البحث، وتتكون هذه البرامج من

الزاحف لعملية الإضافة، وبرنامج المكشف لعملية التحليل والتكشيف، ثم آليات البحث التي تسمح للمستفيد بصياغة وتشكيل استفسارات البحث. وفي حالة محركات بحث الويب غير المرئية فإن عملها يعتمد أساسًا على الزاحف ثم البرنامج الوسيط القائم بإدارة عملية البحث وتلقي النتائج. لقد أفرزت بعض الدراسات نماذجَ للبرامج الوسيطة في بحث قواعد البيانات داخل محركات البحث، ومن هذه النماذج برنامج Semaform : ويعمل هذا البرنامج بالارتكاز على قاعدة معرفية لتركيب وفهم استفسارات البحث. ويقوم هذا الوسيط بإدارة مهام محددة هي[1]:

أولاً: قراءة نماذج البحث web search forms في قواعد البيانات وتحديد خصائصها من خلال برنامج الزاحف.

ثانيًا: استقبال استفسارات البحث من المستفيدين عن طريق واجهة البحث العامة

ثالثًا: توجيه استفسارات البحث الواردة من المستفيدين إلى نماذج بحث قواعد البيانات.

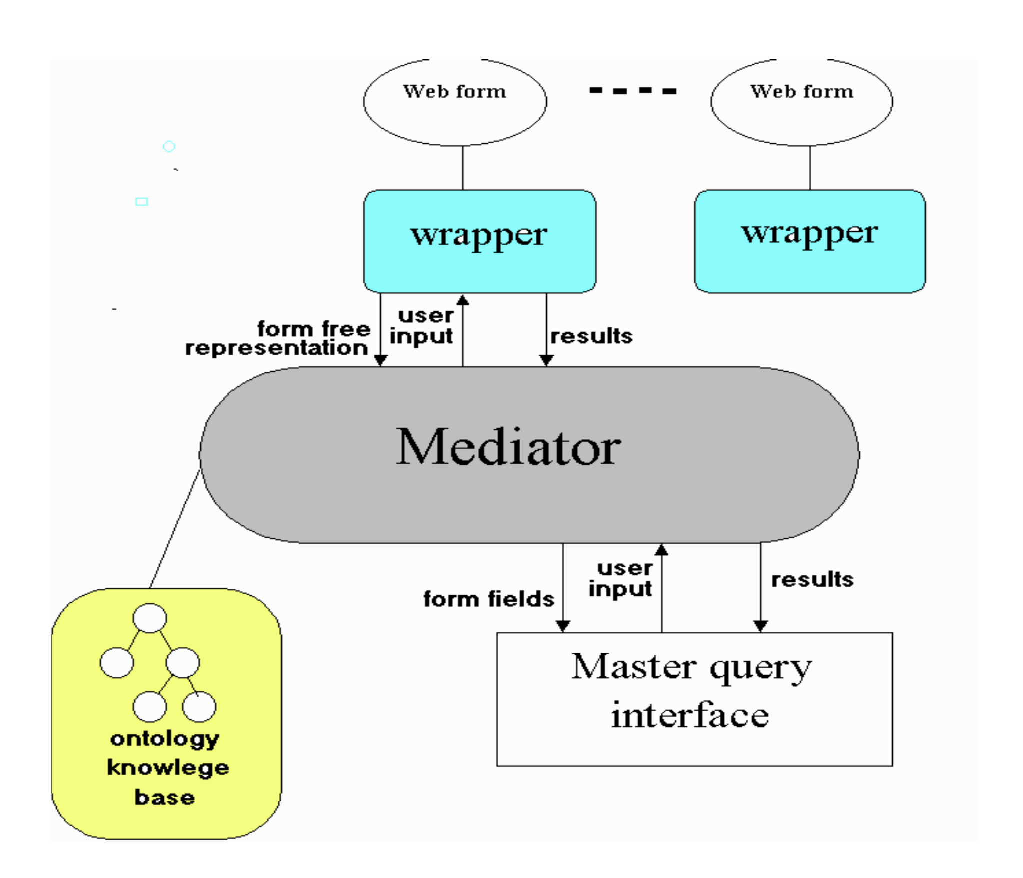
رابعًا: تلقي النتائج المسترجعة وإجراء مختلف عمليات الفرز وإعادة الترتيب.

خامسًا: عرض نتائج البحث المجمعة والمعاد تنظيمها، ثم تقديمها بما يتوافق ومتصفح مستخدم الويب.

ويعمل برنامج semaform في نطاق الويب خارج محركات البحث، وقد عكس ذلك أن هذا البرنامج يعتمد على إدخال وإضافة مواقع الويب موضع البحث من خلال مستخدم الويب وأيضًا من خلال برامج إضافية خاصة به

(١) Walny, Jagoda. SemaForm: Semantic Wrapper Generation for Querying Deep Web Data Sources, Supervisor: Dr. Denilson Barbosa, cited at 10/1/2009, cited at http://www.acm.org/fcrc/PlenaryTalks/Garcia-Molina.pdf

كبرنامج الزحف في محرك البحث. وعلى ذلك فإن مستخدم الويب لا يقوم فقط بإدخال استفسار البحث وإنما يستبق ذلك إدخال نموذج الويب أو قاعدة البيانات المراد البحث بها. ويقوم برنامج الوسيط اعتمادًا على قاعدة المعرفة الموجودة داخله بقراءة نموذج الويب وتحليله لصياغة نمط البحث المتوافق معه من خلال أدوات (wrappers). ويوضح الشكل (٦-٨) أداء مكونات برنامج semaform.

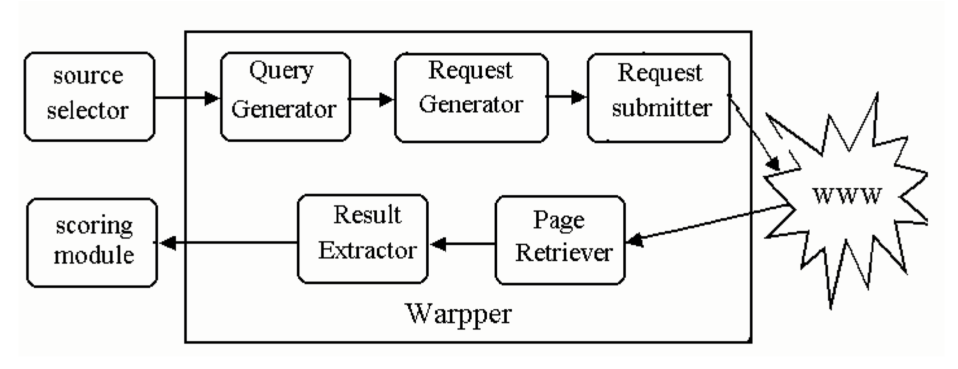


الشكل (٦-٨) أداء مكونات محرك semaform **لبحث نماذج الويب غير المرئية**

يوضح الشكل (٦-٨) المكونات العامة لنظام semaform والمهام التي تضطلع بأدائها، وهي كالتالي :

☒ أولاً: واجهة التفاعل مع الويب the web form interface (wrappers):

تتعامل واجهة الويب مبشرة مع قواعد بيانات الويب : حيث تؤدي مهامَّ مختلفة تبدأ بقراءة أسماء وعناوين فراغات البحث search boxs labels، إلى جانب ملء هذه النماذج باستفسار البحث الموجه لقاعدة البيانات، وأخيرًا تلقي النتائج من قواعد البيانات وتحويلها إلى واجهة البحث الرئيسية أمام مستخدمي الويب. وتأخذ واجهة الويب مكانًا وسطًا بين الوسيط ونماذج بحث واجهات قواعد البيانات. وتعتمد في تكوينها على عنصرين أساسيين يوضحهما الشكل (٦-٩).



الشكل (٦-٩) المكونات الداخلية لواجهة التفاعل مع قواعد البيانات

العنصر الأول : قارئ عناوين النماذج the Label Reader : ويقوم هذا الجزء بقراءة عناوين نماذج البحث لتكوين مخطط أو عرض شجري لواجهة البحث، ثم حفظ عناوين فراغات البحث بحيث يكون كل منها عقدة في شجرة العرض العام لواجهة البحث. بالإضافة إلى حفظ قيم المحتوى الداخلة في نموذج البحث مع حفظ الهيكل البنائي structural information للمعلومات الداخلة في نموذج البحث. ويستطيع قارئ النموذج تحديد الهيكل البنائي للنموذج من خلال تفحص مجموعات الحقول المترابطة معًا. وملاحظة الفواصل والمسافات بين مجموعات النموذج العام، أيضًا الحقول البديلة والمترابطة، مثل حقول البحث العامة والحقول التابعة لها. ومن ثم يكون لكل حقل بحث أساسي وفرعي تمثيل داخل بناء نموذج البحث.

العنصر الثاني : مسترجع النتائج Results Retriever : يقوم مسترجع النتائج بملء نموذج بحث الويب بمجموعة البيانات الواردة له من الوسيط، التي تم إدخالها من قبل بواسطة مستخدم محرك البحث. ويتم ملء البيانات في أماكنها الصحيحة بالاستعانة بالهيكل أو المخطط المعد مسبقاً بواسطة قارئ النموذج. ويأتي دور مسترجع النتائج في استقبال النتائج بعد طرح استفسار البحث، ومن ثم يوجه هذه النتائج إلى الوسيط مرة أخرى كي يتم معالجتها ثم عرضها على متصفح مستخدم الويب.

☒ **ثانيًا: قاعدة المعرفة the knowledge base**:

تحتوي قاعدة المعرفة في محرك البحث المقترح (STORM) على نوعين من المعلومات ؛ يختص **الأول** بالمعلومات الواردة من مستخدم الويب حول طبيعة موضوع البحث Ontology. ويختص **الثاني** بالمجال الموضوعي الموافق لكل نموذج من واجهات بحث قواعد البيانات domain. وتبنى المصطلحات والكلمات الدالة على موضوعات البحث في قاعدة المعرفة بناءً هرميًا يعكس الروابط المنطقية بين المفاهيم والمصطلحات، وتحوي الويب كثيرًا من الأشكال الهرمية لتركيب موضوعات المعرفة وتقسيمها. وجدير بالذكر أن عرض البناء الهرمي للموضوعات أمام مستخدم الويب يمكنه من اختيار المصطلحات ذات الدلالة والمعلومة لدى قاعدة البيانات، وهذا ما يؤثر بدوره في سهولة إجراء البحث نتيجة للتوافق في المصطلحات بين مستخدم الويب وقاعدة المعرفة. ويمكن القول أن اعتماد قاعدة المعرفة لتقسيم هرمي للموضوعات يعرض أمام مستخدم الويب ينتج عنه ميزات عدة منها :

١) تؤدي قاعدة المعرفة دور الوسيط لتحقيق التوافق بين المصطلحات الواردة في نماذج بحث واجهات قاعدة البيانات، التي ستكون المصطلحات نفسها الواردة في استفسار مستخدم الويب عند اختياره لها من التقسيم أو البناء الهرمي المعروض أمامه، وبين كلمات استفسار البحث.

٢) يحقق التوافق السابق بين مستخدم الويب وواجهات قواعد البيانات إيجابية في سرعة النتائج المسترجعة ودقتها.

٣) يساعد البناء الهرمي على تحقيق التوافق في عمليات البحث والاسترجاع مع بعض البيئات مثل بيئة XML.

وتنمو قاعدة المعرفة حجمًا وكمًا من المصطلحات الواردة بها بتزايد عمليات البحث والوصول إلى نماذج واجهات قواعد بيانات جديدة، حيث تمدها نماذج البحث حديثة الإضافة بمجموعات من المصطلحات الدالة على مجالات موضوعية مختلفة. وهو ما يدعم عمليات المضاهاة بين مصطلحات الاستفسار والنماذج الموجودة داخلها، بالإضافة إلى دعم عمليات تحسين وتنقيح النتائج المسترجعة من قواعد البيانات.

☒ ثالثًا: نموذج البحث الرئيسي the master query form:

تمثل واجهة البحث لنظام الاسترجاع نموذج البحث الرئيسي، التي يتعامل من خلالها مستخدم الويب مع النظام كله، طارحًا استفساره ومتلقيًا نتائج بحثه. وتتكون هذه الواجهة من مجموعة من حقول البحث تعبر عن النموذج العام الذي يتبناه نظام الاسترجاع لبحث قواعد البيانات. وتعكس الحقول الواردة في واجهة البحث حقول النموذج الأساسي داخل قاعدة المعرفة. وتقوم هذه الواجهة بمهمتين رئيستين هما :

☒ تلقي استفسار البحث Query Interface :

بحيث يضع مستخدم الويب مختلف المصطلحات الدالة على استفساره بحسب تقسيم الحقول المعروضة على واجهة البحث الرئيسة.

☒ عرض النتائج Displaying Results :

بعد الانتهاء من طرح استفسار البحث تطرح كل قاعدة من قواعد البيانات المتوافقة مع موضوع البحث النتائج المتصلة بالموضوع، ثم تمرر هذه النتائج على الوسيط لمعالجتها حتى تعرض بما يتوافق ومتصفح مستخدم الويب.

☒ رابعًا: برنامج الوسيط the mediator :

يمثل برنامج الوسيط قلب نظام الاسترجاع ؛ حيث يقوم بمهام رئيسة تتمثل في إجراء المضاهاة بين نماذج قواعد البيانات والمصطلحات الواردة من مستخدم الويب، إلى جانب تصميم النموذج الرئيس للبحث، وإرسال استفسار البحث إلى قاعدة أو قواعد البحث الموافقة للموضوع ثم تلقي النتائج وترتيبها وتنقيحها لعرضها على واجهة عرض النتائج الرئيسة. وعلى ذلك فإن الوسيط يعد الداعم لمختلف مهام نظام الاسترجاع كله، حيث يستقبل نماذج بحث الويب من واجهة التعامل مع الويب wrappers ، ثم مضاهاتها مع المصطلحات الدالة الموجودة في قاعدة المعرفة وتسكين قواعد البيانات في مجالاتها الموضوعية الملائمة. إلى جانب تلقي استفسار مستخدم الويب عن طريق نموذج البحث العام المصمم عن طريق الوسيط، وإجراء عمليات المضاهاة لإرسال الاستفسار إلى جهته الصحيحة واستقبال النتائج وتنقيحها وعرضها.

إن أهم ما ينبنى عليه عمل الوسيط هو عملية المضاهاة التي تتسم بمجموعة ملامح يجب مراعاتها مثل:

- **التشابه بين المصطلحات في عملية المضاهاة** : حيث يجب أن تعامل المصطلحات المترابطة المتشابهة تعاملا واحدًا أثناء عملية المضاهاة، مثل مصطلحي adult , adults.
- **المترادفات :** عند مقارنة كلمة ما يجب التنبه إلى إدخال مترادفاتها في عملية المقابلة، حتى يتم استدعاء مختلف وحدات الموضوع على اختلافها اللفظي.
- **المفهوم الشامل للكلمة :** حيث تدخل مصطلحات الأبناء والآباء المتواجدة في نموذج واجهات قواعد البيانات في عملية المضاهاة عند البحث بكلمة الأسرة.

- **بنيــة المـصطلحات وترابطهــا :** حيــث يمكــن اســتخدام مجموعــة مــن المصطلحات المترابطة منطقيًا في استفسار البحث، ومضاهاتها بمجموعة عناصر البحث التي تتفرع من عنصر بحث عام.

٣/٢/٦ واجهة البحث وآلياته:

لا تقل واجهة بحث محرك الويب غير المرئية أهمية عن باقي مكونات المحرك؛ حيث تكمن أهميتها في احتواء النموذج العام لبحث مختلف قواعد بيانات الويب. وتحمل هذه الواجهة دلائل كفاءة أو فشل محرك البحث في استرجاع مجموعات المحتوى غير المرئي ؛ حيث يحكم من خلالها المستفيد على مجموعات النتائج المسترجعة أمامه من أكثر من قاعدة بيانات، فضلاً عن قيامها بتلقي استفسار البحث الذي ينتقل بدوره إلى مختلف نماذج بحث قواعد بيانات الويب ذات الصلة بمحرك البحث. وسوف يتم تناول بعض الخصائص المقترحة لواجهة البحث العامة في الجوانب التالية :

١/٣/٢/٦ شكل واجهة البحث:

تعبر واجهات البحث دائمًا عن طبيعة عمل نظام الاسترجاع، وحسبما يجد المستفيد من مرونة في استخدام هذه الواجهة، بقدر ما يستطيع تحقيق صياغة استفسار البحث في أفضل شكل، ومن ثم استرجاع أدق النتائج من قواعد البيانات. وتنقسم واجهات البحث دائمًا في شكلها بين واجهات البحث البسيطة وبين واجهات البحث المركبة. ويأتي تعدد واجهات البحث لتنوع مهارات مستخدمي الويب في إجراء البحث ؛ حيث قد لا يحتاج كثير من مستخدمي محرك البحث سوى وضع الكلمات المفتاحية الدالة على موضوع البحث، وقد يحتاج بعضهم الآخر إلى آليات وعناصر بحث متقدمة لصياغة استراتيجيات البحث المركبة المعقدة.

٢/٣/٢/٦ أساليب البحث المستخدمة:

كشفت الدراسة التجريبية في الفصل الخامس عن نتيجة مهمة وهي أن أدوات بحث الويب من محركات وأدلة بحث تميل إلى استخدام أسلوبي البحث الحر بالكلمات المفتاحية، وأسلوب البحث بالتصفح الموضوعي. وإنما يرجع ذلك إلى أن قدرات العمل لدى أدوات بحث الويب ما زالت في مراحلها الأولية، ومن ثم فإن هذه الأدوات تحاول العمل باستخدام البحث الحر لكونها محركات بحث، والى جانب ذلك فهي تستخدم البحث بالتصفح لكونه الأسلوب الأيسر والأفضل لكل من أداة البحث والمستفيدين في حصر وترتيب واسترجاع قواعد بيانات ومواقع الويب غير المرئية. ويميل الباحث إلى الجمع بين الأسلوبين معًا في بوابة الويب غير المرئية المقترحة في هذه الدراسة، وهو ليس اتجاهًا خاصًا، وإنما اتجاه عام عملت على استخدامه مختلف محركات وأدلة بحث الويب المرئية مثل yahoo, google. والأحرى بأدوات الويب غير المرئية أن تسعى إلى ذلك، خاصة أنها تتعامل مع محتوى صعب التنظيم والاسترجاع.

٣/٣/٢/٦ التوافق وقواعد الويب غير المرئية:

تمثل واجهة البحث العامة لمحرك بحث الويب غير المرئية النافذة التي ينظر من خلالها المستفيدون إلى كل مواقع وقواعد البيانات المرتبطة بمحرك البحث، ولذلك فإن معيار التوافق بين آليات وعناصر البحث في الواجهة العامة وبين آليات وعناصر البحث في قواعد البيانات لابد أن يكون في أفضل حالاته. ويمكن القول إن هذا التوافق يرتبط ارتباطًا طرديًا بكفاءة استرجاع مجموعات قواعد البيانات. ويأتي التأكيد على جانب التوافق هنا، لأن الويب غير المرئية تمتلك نماذج بحث متنوعة في حقول وعناصر البحث داخلها، ومختلفة في طبيعة المواقع التي تعمل بها ؛ فهناك من نماذج البحث ما يعمل باستخدام حقول الوصف الببليوجرافي ومعاملات المنطق البولياني، وهناك من الحقول ما يقتصر على حقل

النص وإدخال الكلمات المفتاحية، ومن ثم فإن مكونات وعناصر نموذج البحث في الوجهة العامة لابد أن تتوافق تمامًا مع نماذج البحث في قواعد البيانات حتى تسترجع أدق النتائج. ويعبر الشكل (٦-١٠) عن نموذج لمكونات واجهة بحث محرك الويب غير المرئية.



الشكل (٦- ١٠) نموذج لمكونات واجهة بحث محرك الويب غير المرئية

٤/٣/٢/٦ خصائص التفاعل والمساعدة:

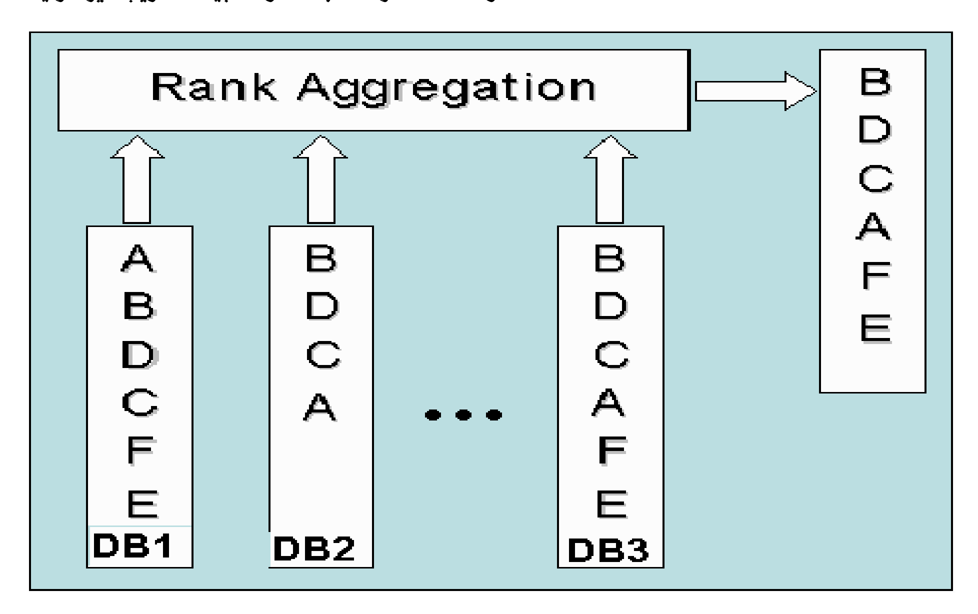
يحتاج الكثير من مستخدمي الويب ذوي مهارات البحث المحدودة إلى التوجيه والمساعدة الدائمة أثناء إجراء عمليات البحث والاسترجاع. ولذلك فإن امتلاك واجهة البحث لرسائل المساعدة وواجهات شرح استخدام محرك البحث هي أهم ما يمكن أن تقدمه للمستفيدين.

٤/٢/٦ ترتيب النتائج المسترجعة وفرزها:

يتلقى محرك بحث الويب غير المرئية مجموعات مختلفة من النتائج المسترجعة من أكثر من قاعدة بيانات، ومن ثم فإن هذه المجموعات من المؤكد أن تتسم فيما بينها بالتداخل أحيانًا والتكرار أحيانًا أخرى. ومن هنا كانت أهمية مرحلة الفرز وإعادة ترتيب مجموعات النتائج مرة أخرى قبل عرضها على واجهة عرض النتائج. وتمثل مرحلة الفرز وإعادة الترتيب نقطة التطابق في الأداء بين محركات بحث الويب غير المرئية وبين محركات البحث المتعددة، لما تقوم به من إعادة ترتيب النتائج الواردة إليها من محركات البحث المختلفة. ومن ثم فإن تناول هذه المرحلة داخل محركات بحث الويب غير المرئية سيكون على مرحلتين متلاحقتين هما:

١/٤/٢/٦ مرحلة الفرز والاستبعاد:

إن التكرار بين المجموعات المسترجعة من قواعد البيانات غالبًا ما يحدث خاصة كلما زاد عدد قواعد البيانات التي يوجه إليها استفسار البحث. ولذلك كانت الحاجة إلى عملية إعادة الفرز واستبعاد المتكرر من هذه النتائج فور ورودها إلى محرك البحث. وما يجعل هذه الخطوة مهمة ودقيقة هو أن أعداد النتائج المسترجعة من أي قاعدة بيانات مهما صغر حجمها سيكون كبيرًا، وهو ما قد يتعارض مع توفير وقت وجهد المستفيد في تصفح أعداد كبيرة من النتائج محدودة الأهمية.



الشكل (٦– ١١) نموذج لفرز النتائج المسترجعة وترتيبها

٦ / ٢ / ٤ / ٢ مرحلة حساب الأهمية وإعادة الترتيب:

بعد أن ينتهي محرك البحث من استبعاد مجموعات النتائج المكررة، تبدأ مرحلة إعادة الترتيب مرة أخرى، وفقًا لأنماط مختلفة من أساليب الترتيب. وقد أشار Lewandowski, Dirk إلى أن هناك الكثير من أنماط ترتيب النتائج المسترجعة داخل نظم استرجاع المعلومات، تنقسم بين نوعين أساسيين يتمثلان فيما يلي[١]:

(١) Lewandowski, Dirk. Web searching, search engines and Information Retrieval, heinrich-Heine-University, 2005, cited at 10/1/2009, cited at http://www.durchdenken.de/lewandowski/doc/isu_preprint.pdf

أولاً: عوامل الترتيب المبنية على استفسار البحث Query-dependent ranking factors:

الجدول (٦- ٢) عوامل الترتيب المبنية على استفسار البحث

مدى تكرار كلمات البحث	عدد مرات تكرار الكلمات المفتاحية في متن الوثيقة
بعد أو صغر المسافة بين كلمات البحث	————
ترتيب كلمات البحث الواردة في الاستفسار	————
موقع كلمات البحث في الوثيقة	حيث تختلف أهمية كلمات البحث بحسب موقعها في الوثيقة، فالكلمات التي تأتي في العنوان أهم من المقدمة، وهي بدورها أهم ما ورد في المتن
Metatags الحقول الوصفية	وهي الكلمات التي ترد في حقول الكلمات المفتاحية
الكلمات ذات النمط المميز	بحيث تكون هذه الكلمات ذات لون أو نمط كتابة مخالف مما يجعلها مميزة بين باقي كلمات المتن.
الكلمات التي ترددت أكثر على مستوى الوحدات ككل	وهي مجموعة الكلمات التي وردت في أكبر عدد من الوحدات المسترجعة ككل
توحيد اللغة	وهي مجموعة الوحدات ذات اللغة الواحدة المتوافقة ولغة واجهة البحث
التغطية الجغرافية	وهو مجموعة الوحدات المتواجدة في مدى قريب للمستفيد بما يجعله يحصل عليها بيسر

ثانيًا: عوامل الترتيب المستقلة عن استفسار البحث Query-independent ranking factors

الجدول (٦- ٣) عوامل الترتيب المستقلة عن استفسار البحث

موضع الوحدة في دليل البحث	حيث يفضل مجموعة الوثائق التي تأتي في بداية الترتيب الهرمي

عـــدد الـــروابط الفائقـــة الـــواردة في الوحدة المسترجعة	كلما زاد عدد الروابط في الوثيقة زادت أهميتها في الموضوع
مدى التردد على الوحدة وشيوع المسار	تقـاس أهميـة الوثيقـة بمـدى تكـرار مسـارها أو روابطهـا في الويب
عدد مرات الزيارة والاستخدام	كلما زاد عدد الاستشهادات زادت أهمية الوثيقة
التحديث	تقارب فترات التحديث
حجم وطول ملف المعلومات	تختلف أهمية الوثيقة بحسب كم المعلومات المتخصصة
نوع وشكل ملف المعلومات	هناك من المستفيدين من يفضل قراءة نوع معين من الملفات as PDF or DOC.
حجـم وعـدد صـفحات موقـع الويـب المسترجع	حجم وطبيعة المعلومات في الموقع مصدر الوثيقة.

ويقترح الباحث أن تتوافر في عملية الفرز وإعادة الترتيب مجموعة الخصائص التالية:

١. أن يقتصر محرك بحث الويب غير المرئية على عدد محدد من النتائج المسترجعة من بين إجمالي النتائج الخاصة بكل قاعدة بيانات، وذلك للتركيز على أهم النتائج المسترجعة فقط.

٢. أن يتحدد معيار الترتيب لكل وحدة مسترجعة بحسب أكبر عدد لمرات تواجدها في مقدمة النتائج المسترجعة من كل قاعدة بيانات، وذلك كما في الشكل (٦-١١).

٣. الاهتمام بعدد الكلمات الدالة في كل وحدة مسترجعة كمعيار أساسي لوزن أهميتها بين النتائج المسترجعة جميعًا.

٤. توحيد معيار حساب الأهمية والترتيب لكل قواعد البيانات، بحيث تتفق جميعًا في ترتيب مجموعاتها المسترجعة بحسب الصلة بالموضوع.

٥. أن يتيح محرك بحث الويب غير المرئية مجموعة أخرى من خيارات الترتيب؛ مثل الترتيب بحداثة المواد، والترتيب بحسب مصدر الاسترجاع. وهذا من أجل أن تتكامل الخدمة المقدمة من جانب محرك البحث أمام المستفيد.

٥/٢/٦ معايير وآليات العمل الداعمة للبحث "Z39.50":

خرج معيار Z39.50 في أول إصداراته عام ١٩٩٢م من خلال المعهد القومي الأمريكي للتقييس والمنظمة العالمية للتوحيد والتقييس ANSI/NISO standard. وهو برتوكول مخصص للعمل مع تطبيقات بحث واسترجاع المعلومات المتنوعة ؛ حيث يسمح بتعيين الإجراءات الضرورية وتنفيذ البحث داخل أنظمة المعلومات المختلفة، ومنها على سبيل المثال قواعد البيانات المختلفة سواء أكانت البيليوجرافية أم غير البيليوجرافية مع إمكانية عرض نتائج الاستفسار في شكل الاتصال العميل أو الخادم، حيث إن كل مستخدم يمتلك برنامج عميل Z39.50 يمكنه إجراء البحث داخل قواعد البيانات المختلفة من خلال خادم Z39.50. ومن المكتبات والمراكز البحثية ومراكز حفظ المعلومات التي قامت باستخدام وتطبيق هذا المعيار نذكر على سبيل المثال مكتبة الكونجرس الأمريكي، والمكتبة الوطنية الكندية[1].

وقد بلغ هذا المعيار من التطور في العمل على الإنترنت مبلغًا يؤهله للدخول ضمن تطبيقات نظم استرجاع قواعد البيانات التي تهدف هذه الدراسة إلى رسم ملامحها، ويتمثل هذا التطور في الملامح التالية:

أولاً: اتخذت كثير من مؤسسات المعلومات قرارها بتنصيب معيار Z39.50 مع برتوكول TCP/IP بدلاً من اندماجه مباشرة مع نموذج OSI. وهناك الكثير من الأسباب التي توضح هذا الاختيار منها على سبيل المثال، أن العمل من خلال بروتوكول TCP/IP هو الأكثر شيوعاً وانتشاراً في الوقت الراهن مقارنة بنموذج

(١) أحمد فرج أحمد. معيار Z39.50 : أداة أساسية لاسترجاع المعلومات عبر بوابات المكتبات .- cybrarians journal، ع ١١ (ديسمبر ٢٠٠٦م) ، تاريخ الاستشهاد ٢٠٠٧/١٠/٢٠م . - مسار : http://www.cybrarians.info/journal/no11/z39.50.htm

OSI، إلى جانب أن التوافق مع TCP/IP يجعل استخدام معيار Z39.50 على شبكة الإنترنت أكثر سهولة.

ثانيًا: لغة الاستفسار (البحث) الشائعة Common Query Language (CQL) : حيث يتم من خلال هذه اللغة تنفيذ البحث داخل أنظمة استرجاع المعلومات المختلفة. وقد صدرت الإصدارة الأولى من هذه اللغة Version 1.1 في فبراير ٢٠٠٤م، وخضعت لبعض التعديلات والإضافات.

ثالثًا: اندماج Z39.50 إلى الشبكة العنكبوتية العالمية (الويب) وذلك بحسب المنهجين التاليين :

المنهج الأول: يرتبط باستخدام خادم ويب مع معيار Z39.50، ويتضمن خادم الويب عميل Z39.50، الذي يستخدم في إعادة صياغة الاستعلامات القادمة من عميل الويب والمتجهة إلى خادم Z39.50 ، وفي العودة يقوم بتنسيق المعلومات القادمة من خادم Z39.50 في شكل HTML ويقوم بإعادته إلى العميل المسئول على عرض نتائج الاستعلام.

المنهج الثاني: يرتبط بعميل ويب مع دعامة من Z39.50 وذلك من خلال عنوان URL، وهنا يقوم عميل الويب بالعمل بالشكل التالي « « ...//:z39.50 » . وتجدر الإشارة إلى أن نظام VTLS يستخدم هذا المنهج.

٦/٢/٦ الخدمات المصاحبة لعمل محرك بحث الويب غير المرئية STORM:

لا يمكن لمحركات بحث الويب أن تعمل كنظم مجردة لبحث واسترجاع المعلومات الرقمية فقط، وإنما تحتاج محركات بحث الويب عامة إلى تقديم مجموعة من الخدمات الداعمة والمكملة لعملية البحث بما يسمح للمستفيد بالاستفادة القصوى من محركات البحث، غير أن هذه الخدمات لا يجب أن تمتد إلى الخدمات الترفيهية ولا يجب أن تنحصر بالدرجة التي ينعدم فيها تواجد هذه الخدمات. وقد عملت الدراسة التجريبية لأدوات بحث عينة الدراسة على الوقوف

على أهم الخدمات المقدمة من محركات وأدلة البحث، وقد خلص الباحث إلى أن أهم هذه الخدمات يتمثل فيما يلي:

١. خدمة تحديد اتجاهات البحث.

٢. خدمة البحث داخل النتائج.

٣. خدمة التواصل مع محرك البحث.

٤. خدمة البريد الإلكتروني.

٥. خدمة إحصاءات حول النتائج.

٦. خدمة بناء سمات المستفيد.

٧. خدمة التوزيع الموضوعي للنتائج.

٣/٦ ملامح بناء أدلة قواعد الويب غير المرئية:

تتسم أدلة بحث الويب بصفة عامة بملامح عمل ذات مرونة عالية، خلاف ما تكون عليه محركات البحث. وقد ساعدت هذه الملامح في أن تكون أدوات بحث الويب غير المرئية في مراحلها الأولى تقترب لكونها أدلة بحث أكثر من كونها محركات بحث فعلية. وتعد ملامح مثل الدقة والبنية الموضوعية الهرمية وارتفاع معامل التحقيق في استرجاع النتائج هي أهم السمات التي يختص بها دليل بحث الويب، فضلاً عن تقديمه للمعلومات الوصفية والإحصائية عن مجموعة مصادر الويب التي تحويها قاعدة بياناته. وقد ذكرت Marsha Luevane أن أدلة بحث الويب عامة هي نظم استرجاع للمعلومات الرقمية، تعتمد في بنائها على أربعة أسس هي[١]:

(١) Luevane, Marsha and Chervenak, Joe. Web Searching 101: Search Tools, Strategies, and Tips, National Renewable Energy Laboratory, 2003, cited at 14/5/2009, cited at http://www-conf.slac.stanford.edu/interlab99/program/chervenak-luevane/searching101.ppt

١) أن أدلة بحث الويب هي أساسًا قوائم حصرية لمواقع أو قواعد بيانات الويب.

٢) أنها دائمًا تعتمد على الترتيب الموضوعي لتنظيم مجموعاتها.

٣) أنها دائمًا تحمل وصفًا لمواقع أو مصادر معلومات الويب.

٤) أنها دائمًا تبنى وتدار اعتمادًا على العنصر البشري.

والى جانب ذلك فإن أدلة البحث تمتلك مجموعة من الخصائص التي يرى الباحث فيها دعمًا لحل مشكلة استرجاع المحتوى غير المرئي أمام مستخدمي الويب ؛ حيث سيسترجع المستفيد مجموعات منتقاة من قواعد بيانات الويب، التي خضعت لعملية التحكيم قبل ضمها إلى مجموعات الدليل. أيضًا فإن عملية استرجاع مجموعات البيانات ستكون في أعلى درجات التحقيق؛ لاقترابها من استخدام اللغة المقيدة في استخدام نظام تصنيف ما أو إحدى قوائم رؤوس الموضوعات، فضلاً عن سهولة التحرك بين المجموعات والترتيب الموضوعي صعودًا وتعمقًا في القوائم الموضوعية وتفريعاتها. وسوف يتم تناول مقترحات الدراسة في بناء أدلة بحث الويب غير المرئية، اعتمادًا على النقاط التالية:

١/٣/٦ الجمع بين المحرك والدليل :

لقد عملت هذه الدراسة منذ البداية على أن يكون دليل البحث جزءاً مكملاً لمحرك البحث في بناء النظام المقترح لاسترجاع الويب غير المرئية، وهذا يرجع إلى عاملين أساسيين هما ؛ الأول : أن دليل البحث قد لا يكفي بمفرده لحل مشكلة استرجاع المحتوى غير المرئي الممثل في تسجيلات وملفات قواعد البيانات، ذلك لأنه في هذه الحالة لن يدعم آليات البحث الحر التي تستطيع التعامل مع نماذج بحث قواعد بيانات الويب وتوجيه استفسارات البحث لها، ثم استرجاع النتائج مرة أخرى. العامل الثاني: هو أن بناء محرك لبحث الويب غير المرئية منفردًا دون الاعتماد على دليل بحث مكمل له، سوف يُفقد المستفيدين مرونة أسلوب التصفح والتعرف إلى مواقع وقواعد بيانات الويب، فضلاً عن أن آليات وعناصر

البحث الحر حينئذ سوف تبلغ من التقدم والتعقيد الدرجة التي لا تناسب مستوى المهارات المحدود لدى كثير من المستفيدين. وأخيرًا، فإن الغاية التي تسعى إليها هذه الدراسة تتمثل في بناء بوابة لبحث محتوى الويب غير المرئية تعتمد في بنيتها على استخدام المحرك والدليل معاً.

٢/٣/٦ عناصر وصف قواعد البيانات:

لا يقدم دليل البحث ترتيبًا موضوعيًا لمجموعات مواقع ومصادر الويب التي يحويها داخل قاعدة بياناته فقط، وإنما يقدم أيضًا المعلومات الوصفية والإحصائية للتعريف بطبيعة وماهية موقع أو قاعدة بيانات الويب. وتتوزع هذه المعلومات الوصفية على مجموعة ثابتة من الحقول، تقترب أكثر ما تكون إلى تسجيلات ببليوجرافية لوصف مواقع الويب. وقد ناقشت هذه الدراسة في فصلها الثالث عناصر تكشيف ووصف ملائمة للتعريف بقواعد البيانات، ونقدمها هنا كتسجيلة مقترحة لوصف قواعد البيانات داخل نظام استرجاع قواعد بيانات الويب غير المرئية. وقد ضم الجدول (٦-١) السابق هذه حقول الوصف المقترحة وهي :

- عنوان قاعدة البيانات.
- الوكيل أو المالك.
- مستخلص عن قاعدة البيانات.
- دورية وفترات التحديث لقاعدة البيانات.
- الإتاحة الشرعية للوصول لقاعدة البيانات.
- الأشكال الرقمية لقاعدة البيانات.
- تكاليف النسخ أو مقابل التحميل.
- تكاليف إعادة الإنتاج لقاعدة البيانات.

- ❖ تقرير رسوم خدمات قاعدة البيانات.
- ❖ قائمة نماذج البيانات داخل قاعدة البيانات.
- ❖ قائمة حقول التسجيلات داخل قاعدة البيانات

٣/٣/٦ آلية التقسيم الموضوعي وتصفح قواعد البيانات:

يعتمد دليل البحث على مبدأ التقسيم الموضوعي للمجالات العامة أو المجالات المتخصصة التي يشملها دليل البحث. ويختلف مستوى التقسيم الموضوعي داخل دليل البحث من المستوى البسيط إلى المستوى المعقد، وهو ما يعكسه عدد قوائم الموضوعات الرئيسية والفرعية داخل دليل البحث ؛ فكلما زاد عدد هذه القوائم، زادت درجة تعمق دليل البحث في المجال الموضوعي، والعكس صحيح. وإلى جانب ذلك فإن أدلة بحث الويب تختلف فيما بينها في الاعتماد على نسق الترتيب المنطقي المنهجي للتفريعات الموضوعية أو الاعتماد على تقسيم عام للموضوعات يخلو تمامًا من أي ترتيب منطقي أو هجائي لمصطلحات وموضوعات قوائم دليل البحث.

وقد كشفت الدراسة التجريبية لعينة أدلة بحث الويب غير المرئية، عن أن أدلة الويب غير المرئية جميعها مازال يفتقر إلى اتباع نظام ترتيب منطقي كخطط التصنيف أو نظام ترتيب هجائي كقوائم رؤوس الموضوعات، واتفقت مفردات عينة أدلة البحث على العمل بتقسيم عام للمصطلحات الموضوعية، غير متبعةٍ في ذلك أياً من علاقات الربط المنطقي أو الهجائي بين المصطلحات. ويضاف لهذه الحقائق أن التقسيمات الموضوعية داخل أدلة البحث لم تتعدَّ تفريعاتها المستوى الثاني أو الثالث في أفضل الحالات، وتمثلت أفضل التقسيمات الموضوعية في دليل البحث completplanet. وتعرض الدراسة هنا مقترحاتها في بناء التقسيم الموضوعي لدليل بحث الويب غير المرئية، وفقًا للجوانب التالية:

١/٣/٣/٦ أنماط التقسيم الموضوعي لأدلة البحث:

يمكن لكثير من أدلة بحث الويب أن تتوافق معًا في اتباع نسق أو نظام ترتيب واحد، غير أنها قد تختلف في نمط التقسيم الموضوعي وشكله. والمقصود هنا بنمط الترتيب هو عنصر الترتيب داخل قوائم الموضوعات والتقسيمات الموضوعية ؛ فيوجد الترتيب بالمصطلحات والترتيب بالوحدات المصورة ... إلخ. ويأتي تعدد أنماط الترتيب من تعدد أشكال التعبير عن الموضوعات داخل دليل البحث، فضلاً عن أن شكل وطبيعة مصادر معلومات دليل البحث قد تفرض اتباع نمط محدد دون غيره، مثل أدلة بحث النصوص والصوت والصورة. وتتمثل أهم أنماط التقسيم الموضوعي فيما يلي:

- ❖ التقسيم الموضوعي بالمصطلحات : وهو يتلاءم أكثر مع مجموعات الوحدات النصية مثل المقالات أو الملفات النصية داخل قواعد البيانات.
- ❖ التقسيم الموضوعي باللقطات المصغرة : وهو يتلاءم أكثر مع الصور المتحركة والثابتة.
- ❖ التقسيم الموضوعي بعناوين الوحدات أو المصادر : وهو يتلاءم أكثر مع ترتيب وتنظيم مصادر أو نظم الاسترجاع مثل قواعد البيانات، حيث يكون الاسترجاع هنا على مستوى النظام وليس مستوى المقالة أو الملف.
- ❖ التقسيم الموضوعي بالروابط الفائقة : وهو يتلاءم أكثر مع مواقع وصفحات الويب.

٢/٣/٣/٦ نظم التقسيم الموضوعي لأدلة البحث:

تحتوي نظم استرجاع المعلومات التقليدية والرقمية على خطط تصنيف وقوائم رؤوس الموضوعات، لإتمام عمليات التحليل الموضوعي لمصادر المعلومات داخل هذه النظم. ومن ثم فإن خطط التصنيف هذه هي دعامة البناء والتنظيم الموضوعي في

نظم استرجاع المعلومات الرقمية على الويب. وتنقسم نظم التقسيم الموضوعي عامة إلى قسمين رئيسين هما :

☒ **نظم التقسيم المنطقي لرؤوس الموضوعات:**

١. نظام تصنيف ديوي العشري.

٢. نظم التصنيف العشري العالمي.

٣. نظام تصنيف مكتبة الكونجرس.

☒ **نظم الترتيب الهجائي لرؤوس الموضوعات:**

١. نظام قائمة رؤوس الموضوعات العربية.

٢. نظام قائمة سيرز.

٣. نظام قائمة مكتبة الكونجرس.

ويرى الباحث أن اعتماد أحد نظم الترتيب المنطقية مثل نظام تصنيف ديوي العشري أو نظام الكونجرس هو الأمثل في بناء دليل بحث قواعد البيانات ؛ لأنه يتدرج من الموضوعات العامة إلى الموضوعات المتخصصة في شكل منطقي، بالإضافة إلى إمكانية استخدام تفريعاته الدقيقة المتخصصة لترتيب وتصنيف المقالات والمجموعات الداخلية لقواعد بيانات الويب.

٦ / ٤ قضايا بحث قواعد بيانات الويب غير المرئية :

إن مشكلة استرجاع المحتوى غير المرئي لمواقع وقواعد بيانات الويب تحتوي على جوانب عدة، قد لا تشكل آليات البحث الطرف أو الجانب الأكبر فيها، وإنما توجد جوانب أخرى كثيرة ترتبط ارتباطًا وثيقًا بحل مشكلة استرجاع مجموعات قواعد البيانات والتغلب عليها. ويحاول الباحث هنا التعرض لأهم هذه الجوانب محاولاً وضع اقتراحات للتغلب عليها، وهي:

٦/٤/١ قضية الاشتراك والمقابل المادي:

لم تكن مشكلة أدوات بحث الويب في استرجاع المحتوى غير المرئي لقواعد البيانات تقتصر على افتقارها لآليات وتقنيات البحث اللازمة لذلك فقط، وإنما واجهت هذه الأدوات عوائق عدة، وهذا بدوره يرجع إلى اختلاف قواعد بيانات الويب فيما بينها في خصائص وآليات العمل. وكان أهم هذه العوائق هو الجانب التجاري لبعض قواعد البيانات التي تفرض على المستفيدين تسديد قيمة كاملة أو اشتراك دوري مقابل الحصول على خدمات وملفات قاعدة البيانات. ومن ثم مثلت الرسوم المادية حاجزًا كبيرًا بين محركات البحث ومحتوى قواعد البيانات.

ويقترح الباحث في هذا الشأن أن يتم إبرام سياسات أو اتفاقات وتراخيص تعطي محركات بحث الويب حق الوصول إلى مجموعات قاعدة البيانات، واسترجاعها أمام المستفيدين. وفي حين يتم ذلك، فإن لمحرك البحث الحق في اتخاذ إجراءات مختلفة لتعويض هذه النفقات باتباع التالي:

١. أن يقدم محرك البحث بعض خدماته بمقابل مادي.

٢. يمكن لقواعد البيانات تقديم إتاحة واسترجاع ذي مستوى مختصر من التسجيلات، في حين تحتفظ بالمعلومات الكاملة تسترجعها لحين وصول المستفيد إليها بشكل مباشر.

٦/٤/٢ قضية حماية الملكية لقواعد البيانات:

تحمل كثير من قواعد البيانات بعض أشكال المعلومات التي تملك وحدها حق نشرها أو بثها للمستفيدين، ومن ثم فإنها تتخذ مختلف التدابير لحماية هذه الملكية. وحتى يمكن لمحرك بحث قواعد البيانات استرجاع هذه المعلومات مع الحفاظ على الملكية الفكرية لقواعد البيانات، فإنه يمكن لقواعد البيانات اتباع مختلف التقنيات الحديثة في حماية معلوماتها ؛ مثل وضع العلامات المائية

علـى هـذه المعلومـات، أو أن يـتم تقـديم نمـوذج المعلومـات المرئـي في أدنـى درجـات الكثافة له.

٣/٤/٦ قضية أمن المعلومات السرية:

تتنوع قواعد بيانات الويب لتشمل قواعد البيانات التجارية وقواعد البيانات الشخصية وقواعد البيانات العامة وقواعد البيانات الرسمية أو الحكومية، وهذا النوع الأخير من قواعد بيانات الويب قد تجد محركات البحث صعوبة في التعامل مع مجموعاته، إضافة إلى ما تفرضه قواعد البيانات ذاتها من حظر الوصول أو استخدام الملفات الداخلية. وقد لا تحتاج هذه القواعد بذاتها إلى اقتراحات للعمل، لأنها بطبيعتها وطبيعة تواجدها على الإنترنت فإنها حتمًا ما تتخذ الإجراءات الوقائية ضد اختراق البرامج أو الأشخاص لمحتواها والتعرف إليه. وإن كان في هذا الأمر صعوبة كبيرة، خاصة إذا ما تذكرنا وقائع عدة في اختراق قواعد البيانات البنكية والعسكرية. وهنا يمكن القول إن برنامج الزاحف ذاته يستطيع الابتعاد التلقائي عن هذا النوع من القواعد، وذلك إذا ما وضع في ملف الزاحف robot.txt شرط تجنب مسارات الويب ذات الامتداد gov, mil.

٤/٤/٦ قضايا ترشيح المعلومات:

إن محرك البحث المقترح في هذه الدراسة STORM لا يتوقف عند قواعد البيانات النصية وحسب، وإنما يمتد اهتمامه أيضًا إلى استرجاع محتوى قواعد بيانات الوسائط المتعددة، ولهذا كانت قواعد بيانات الوسائط جزءاً من عينة الدراسة من قواعد البيانات. وتبعًا لأن قواعد بيانات الوسائط المتعددة قد تحمل موادَّ من صور أو أفلام يتنافى استرجاعها مع الثقافة العربية الإسلامية، فإن الباحث يرى ضرورة صياغة مختلف تعليمات الترشيح الخاصة ببرنامج الزاحف، فضلا عن أن يحتوى محرك البحث ذاته على برنامج لترشيح المواد المسترجعة اعتمادًا على آليات الترشيح بالنص والترشيح بالمحتوى.

٦/٥ خاتمة :

عملت الدراسة في هذا الفصل على تقديم مواصفات مقترحة لمحرك بحث الويب غير المرئية STORM، وذلك من خلال أربعة محاور ؛ يختص المحور الأول منها بطبيعة تكشيف وتحليل قواعد بيانات الويب، وآلية الوصول إلى نماذج واجهات البحث وتكشيفها من جانب الزاحف. أما المحور الثاني فقد عمل على تحديد الملامح المثلى التي يجب أن يبنى عليها برنامج الزاحف داخل محرك بحث الويب غير المرئية، وآليات عمله الداخلية في الوصول إلى قواعد البيانات وطرح الاستفسارات الأولية عليها، ثم تلقي النتائج المسترجعة وتحليلها لتحديد قرار ضم أو إهمال نموذج بحث قاعدة البيانات. وتمثل المحور الثالث في دراسة خصائص أدلة البحث كجزء داعم ومكمل لآلية البحث الحر واليات البحث داخل نظام استرجاع الويب غير المرئية، وتم تناول الأدلة بالتركيز على ماهية التقسيمات الموضوعية من حيث الأنواع والأنماط والملامح المثلى لبناء هذه الأدلة. وجاء المحور الرابع للتعبير عن الجوانب الأخرى المتعلقة باسترجاع المحتوى غير المرئي، التي تختلف عن عوائق البحث وبناء المحركات، وتمثلت هذه الجوانب في قضايا الملكية الفكرية، وحماية أمن وسرية المعلومات، والتأثر بالعمل في ضوء الثقافة العربية الإسلامية.

وعلى ذلك يمكن القول إن هذا الفصل قد ضم مختلف جوانب الوصف المقترحة اللازمة لبناء محركات بحث واسترجاع المحتوى غير المرئي، وقد سُمي محرك بحث الويب المقترح هنا باسم (STORM)، وقد شرح الفصل الثالث اختيار الباحث البوابات أداة بحث للتعبير عن الأداة النموذجية في بحث قواعد البيانات، لكونها ينطبق عليها مختلف المواصفات اللازمة لبناء نظم استرجاع الويب غير المرئية.

النتائج والتوصيات

أولاً: النتائج:

خرجت الدراسة بمجموعة من النتائج تتمثل في الجوانب التالية:

١. إن الويب تملك كمًا هائلاً من مجموعات المحتوى غير المرئي، غير أن هذا المحتوى ما زال قيد الدراسة للتعرف إلى خصائصه والانتباه لأهميته. ويعكس ذلك أن مصطلحات الدلالة على الويب غير المرئية لم تزل غير مستقرة، وهناك مصطلحات متعددة مثل؛ الويب العميقة deep web، والويب الخفية hidden web، والويب غير المرئية invisible web، الويب الخاصة private web، الويب غير الشفافة opaque web تأتي جميعًا في أدبيات الويب غير المرئية كمترادفات تحمل معنى واحدًا، في حين أنها لا تتساوى في المعنى، وإنما يعبر كل منها عن مستوى معين من مستويات الوصول إلى محتوى الويب غير المرئية. ويفضل الباحث مصطلح الويب غير المرئية، ويرى تعميم العمل به.

٢. ما زالت محركات وأدلة بحث الويب غير المرئية لا تملك حتى الآن القدرات الدلالية والاصطلاحية للتعريف بها، حيث كشفت عينة الدراسة قدرًا كبيرًا من التداخل بين مجموعات محركات وأدلة بحث الويب غير المرئية. هذا فضلاً عن أن هذه الأدوات لا تملك القدرات الفعلية لتنظيم واسترجاع محتوى الويب غير المرئية، مما يعكس أن هذه الأدوات كانت بحاجة إلى دراسة نظرية وتطبيقية على واقع العمل بها، لكشف نقاط القوة والضعف في آليات وبرامج عملها، وهو ما قام به الباحث محاولاً استنباط الخصائص المقترحة لعمل أدوات بحث الويب غير المرئية.

٣. إن الاختلاف بين عمل محركات بحث الويب المرئية وبين محركات بحث الويب غير المرئية يكمن في تغيير نمط وآلية عمل برنامج الزاحف، الذي سيقع عليه عبء الكشف عن نماذج بحث واجهات مواقع وقواعد البيانات، ومن ثم تحليلها وتكشيفها للوقوف على عناصر واليات البحث داخل هذه النماذج.

وهذا يختلف عن عمل برنامج الزاحف في الويب المرئية الذي يتوقف عمله على الوصول إلى صفحات ومواقع بنية html.

٤. تتمتع أدلة بحث الويب غير المرئية بالوفرة العددية وإمكانات البحث والاسترجاع للمحتوى غير المرئي ؛ حيث كان من الصعب على الباحث أثناء تجميعه لعينة الدراسة ومسح أدوات البحث أن يجد محركات لبحث الويب غير المرئية بقدر ما وجد أدلة تحاول تنظيم واسترجاع المحتوى المرئي. ويرجع الباحث ذلك إلى أنه من اليسير على مصممي نظم استرجاع الويب بناء قوائم وتقسيمات موضوعية لمواقع وقواعد بيانات الويب غير المرئية، بينما من الصعب عليهم بناء برامج الزاحف التي تستطيع اختراق محتوى قواعد البيانات، وهذه البرامج تمثل بذاتها رأس حربة العمل لمحركات البحث.

٥. كشفت الدراسة عن صحة الفروض التي بنيت عليها من البداية، ومن ثم فإن الباحث يرى الاعتماد عليها في مجال الويب غير المرئية ؛ حيث إنها الآن تمثل أولى النظريات الخاصة ببناء محركات بحث الويب غير المرئية. وهذه النظريات كما يلي:

أ- لا تستطيع محركات بحث الويب بنمطها الحالي التعامل مع قواعد بيانات الويب ذات الأشكال المختلفة من مصادر المعلومات.

ب- إن اختلاف طبيعة بناء قواعد البيانات (الويب غير المرئية) عن صفحات الويب html (الويب المرئية) يحتاج بدوره إلى جيل متطور من محركات البحث بمعايير أكثر تطور.

٦. تنتهج مختلف أدوات بحث الويب الحالية منهجًا واحدًا في استرجاع المحتوى غير المرئي، وهذا المنهج مؤداه أن تحصر أدوات البحث مختلف مواقع الويب وقواعد البيانات ثم تعمل على وضع عناوينها في قوائم مرتبة هجائيًا كروابط للمواقع ذاتها، غير أن المطلوب فعليًا هو أن تقوم أدوات البحث بالتعرف إلى

محتوى مواقع الويب غير المرئية وتقسيمها إلى مجالات موضوعية عامة، ثم موضوعات أكثر تخصصًا وصولاً إلى المواقع ذاتها، وما زال الدخول إلى محتوى قواعد بيانات الويب غير المرئية يحتاج إلى مزيد من الاهتمام والتطوير. ومن ثم كانت أدلة بحث الويب غير المرئية هي الأكثر انتشارًا واستخدامًا من محركات البحث.

٧. إن تنظيم واسترجاع المحتوى غير المرئي على الويب لا يحتاج إلى محرك بحث منفرد أو إلى دليل بحث يعمل بشكل مستقل، وإنما كشفت الدراسة عن ضرورة العمل على دمج أداتي محرك ودليل البحث في أداة واحدة تستخدم لاسترجاع معلومات الويب غير المرئية. وهذه الأداة ستكون بوابات الويب غير المرئية invisible web portals ؛ حيث ستجمع هذه البوابة بين عمل المحرك والدليل معاً، فضلاً عن التزود بمجموعة من الخدمات والخصائص وآليات العمل التي من شأنها التكامل نحو تنظيم واسترجاع المحتوى غير المرئي على الويب.

٨. إن بداية العمل على تنظيم واسترجاع محتوى الويب غير المرئية، تشبه إلى حد كبير بداية التعامل مع الويب المرئية من مواقع وصفحات. ويمتد هذا التشابه إلى بداية كل من أدوات بحث الويب المرئية مع غير المرئية ؛ حيث عكست أدوات بحث عينة الدراسة أن جملة الاهتمام تنصب الآن على تنظيم واسترجاع المحتوى النصي غير المرئي، مع الإهمال الجزئي للوسائط المتعددة داخل الويب غير المرئية. فقد بدأت محركات بحث الويب بالعمل على تنظيم وبحث الصفحات والمواقع التي تحوي المعلومات النصية، ثم خرج من بعد ذلك جيل جديد من محركات البحث يهتم بتنظيم ومعالجة المواد المصورة والمسموعة.

٩. تواجه أدوات بحث الويب غير المرئية مشكلة استرجاع المحتوى غير المرئي على مستويات ثلاثة مختلفة في درجة التعمق، هي:

الأول : مستوى استرجاع الملفات، وهي أعمق نقطة يمكن أن تصل إليها أدوات بحث الويب غير المرئية ومن ثم فهو أصعب المستويات مشكلة استرجاع المحتوى غير المرئي ؛ حيث تحوي الملفات النهائية غاية المعلومات التي يحتاجها المستفيد، مثل النصوص الكاملة أو الصور الرقمية المتحركة أو الثابتة، غير أن هذه الملفات لا تنتمي إلى قواعد بيانات الويب فقط، بل تتعدى ذلك إلى مواقع الويب التفاعلية.

المستوى الثاني : مستوى استرجاع مواقع الويب، ويعني اكتفاء أدوات البحث بعرض عناوين مواقع الويب غير المرئية، حتى يتمكن المستفيد من الوصول إليها والبحث بها مباشرة، وهو دور الوسيط فقط أو الرابط بين المستفيد ومواقع الويب غير المرئية.

المستوى الثالث : مستوى استرجاع محتوى قواعد البيانات، وفي هذا المستوى تعمل أدوات البحث على اختراق أبواب قواعد البيانات واسترجاع محتواها من التسجيلات الداخلية أو تخطي ذلك إلى المستوى الأول وهو استرجاع الملفات النهائية داخل قاعدة البيانات. ومن ثم فأمام محركات البحث الاختيار بين مستويات العمل المختلفة.

١٠. يتم تحليل وتكشيف محتوى قاعدة البيانات على مستويين يمكن اتباعهما عند بناء أدلة بحث قواعد بيانات الويب غير المرئية ؛ أما المستوى الأول فيتمثل في التكشيف أو التحليل العام : وهنا يتم استخدام المصطلحات الموضوعية والكلمات الكشفية للتعبير عن المجال أو التخصص الموضوعي لقاعدة البيانات كلها، وليس مجموعاتها أو مقالاتها الداخلية. ويستخدم هذا المستوى في دليل البحث حتى يتم إدراج عناوين قواعد البيانات أسفل المصطلح أو المصطلحات الموضوعية في قوائم التصفح التابعة لدليل البحث، ومن ثم يتمكن المستفيد من الوصول لقواعد البيانات الملائمة إذا ما اختار مصطلحًا

أو أكثر من هذه المصطلحات. أما المستوى الثاني فهو التكشيف الداخلي لمجموعات قاعدة البيانات : وهنا يتم تكشيف وتحليل محتوى كل مقالة أو وحدة داخلية داخل قاعدة البيانات، والهدف من ذلك هو أن يتمكن المستفيد من الوصول إلى استرجاع المقالات أو الوحدات الداخلية دون الوقوف عند قاعدة البيانات ذاتها، ويصلح هذا المستوى أكثر في حالة اعتماد دليل البحث على أسلوب البحث الحر إلى جانب قوائم التصفح بالمصطلحات أو الكلمات الدالة.

١١. إن طبيعة بناء قواعد البيانات قد أثرت في درجة اهتمامها باستخدام آليات البحث ؛ حيث جاءت القواعد ذات الطابع التجاري كأكثر القواعد احتواءً على آليات البحث ؛ مثل قاعدتي amazon في القواعد النصية، وقاعدة corbis في قواعد الوسائط المتعددة. ثم جاءت قاعدة eric المتخصصة في مجال التربية كحلقة وسطى بين القواعد الترفيهية والقواعد التجارية. وهذا يعني بشكل عام أننا كلما اتجهنا من المجالات العلمية المتخصصة أو الطبيعة التجارية الخالصة إلى الجوانب الترفيهية، فإننا نفقد الاهتمام باستخدام آليات البحث شيئًا فشيئًا.

١٢. إن قواعد بيانات الويب غالبًا ما تحمل محتوى يختلف بدوره عن طبيعة صفحات ومواقع الويب، وذلك لما لهذا المحتوى من خصائص تتنوع بين دقة الصياغة ودقة البناء وشدة التقييد في الضبط والتحكم ؛ حيث تبعد أكثر ما تبعد قواعد البيانات عن المحتوى الغث الشائع في مواقع الويب الأخرى. وعلى ذلك فإن المحتوى الذي تقدمه هذه القواعد دائمًا ما يشترط للوصول إليه المقابل المادي أو الاشتراك أو الانتماء إلى هيئات علمية يمكن للمستخدم من خلالها الحصول على محتوى قواعد البيانات. وتعد هذه الشروط بمثابة درجات تختلف فيها قواعد البيانات ذات النشاط التجاري عن قواعد البيانات

العلمية المتخصصة إلى قواعد البيانات الترفيهية التي قد لا تفرض أي شروط للوصول إلى محتوياتها.

١٣. تمثل قضية مستوى الوصول إلى محتوى قواعد بيانات الويب حجر عثرة أمام محاولات تكشيف هذه القواعد وإتاحتها، لأن الكثير من قواعد البيانات لا يمكن النفاذ إلى محتواها دون الرسوم المادية، كما أن هناك قواعد البيانات المحاطة بسرية وأمن المعلومات. ومن ثم فإن صياغة نموذج من اتفاقيات أو إجازات الوصول licenses إلى المحتوى يتم إبرامها بين محركات بحث الويب غير المرئية وبين قواعد البيانات، يحمل من خلال برنامج الزاحف مفتاح الباب المغلق لمحتوى قواعد البيانات التجارية.

١٤. وضعت قواعد بيانات العينة درجتين من درجات الحصول على محتواها الداخلي؛ هما المستوى المفتوح حيث يمكن للمستفيد الوصول إلى محتوى التسجيلات الداخلية لقاعدة البيانات والاطلاع على الملف الأصلي سواء كان النص الكامل أو ملف وسائط متعددة، ومستوى آخر هو المستوى المحدد الذي يقف عند تمكين المستفيد من إجراء البحث والوصول إلى مكونات التسجيلة النهائية، غير أنه لا يمكن الاطلاع على ملفات النصوص أو الوسائط المتعددة التي توجد داخل تسجيلات قواعد البيانات.

١٥. لقد اتفقت قواعد البيانات النصية وقواعد الوسائط المتعددة فيما يتبعونه من أنماط ترتيب النتائج، حيث اعتمدت أغلب قواعد بيانات العينة في شقيها النصية وقواعد الوسائط على مبدأ الترتيب بحسب الصلة بالموضوع subject relevance كمبدأ أساسي لترتيب مجموعة النتائج المسترجعة من قاعدة البيانات. وقد خالف ذلك في كل من القواعد النصية وقواعد الوسائط قواعد البيانات eric, amazon, corbis في كونها قد زادت على مبدأ الترتيب بحسب

الصلة أنماطًا أخرى مثل ؛ الترتيب وفقًا للسعر في amazon، والترتيب وفقًا لهجائية المؤلف أو العنوان في eric، والترتيب وفقًا لتواريخ التقاط الصورة في corbis.

١٦. إن نجاح هذه الدراسة في الوقوف على أهم آليات وملامح عمل محركات وأدلة بحث الويب غير المرئية، هو بمثابة مفتاح بوابة الدخول إلى قواعد بيانات المكتبات الرقمية، ومن ثم فهو آلية جديدة يمكن لقواعد بيانات المكتبات الرقمية من فهارس ونصوص كاملة أن تنخرط بين مصادر معلومات الويب، وتمتزج معها محققة لمستخدمي المكتبات الرقمية إمكانية التعامل معها واسترجاع محتواها من خلال نظام استرجاع الويب غير المرئية.

١٧. تشابه الدور الذي تؤديه واجهة محرك بحث قواعد بيانات الويب غير المرئية مع الدور الذي تقوم به واجهة محرك البحث المتعدد ؛ حيث تستقبل واجهة محرك البحث المتعدد كلمات الاستفسار ثم توجهها إلى محركات البحث ذات الصلة، ثم تتهيأ مستقبلة مجموعات النتائج التي استرجعتها كل محركات البحث. غير أن واجهة محرك البحث المقترح STORM في هذه الدراسة سوف تتعامل مع واجهات بحث قواعد بيانات الويب التي تملك آليات بحث تفوق تلك التي تعتمد عليها محركات بحث الويب، فضلاً عن أن قواعد بيانات الويب تعتمد على استخدام بنية التسجيلات الببليوجرافية التي يمكن أن ينتج عنها حقولاً نشطة للبحث توضع في واجهة قاعدة البيانات حتى يتم استخدامها في استرجاع التسجيلات النهائية.

١٨. إن عمل نموذج بحث محرك قواعد البيانات لا يقتصر على تلقي الاستفسارات البحثية من مستخدمي محرك البحث، وإنما يمتد إلى التفاعل والتعامل مع نماذج بحث قواعد بيانات الويب ؛ حيث يقوم النموذج العام الموجود في واجهة بحث البحث بإرسال استفسار البحث إلى قاعدة أو قواعد بيانات الويب المتوافقة وموضوع الاستفسار. ومن ثم فإن ثمة رابط لا ينقطع يجمع بين نموذج

البحث في المحرك وبين نماذج البحث في قواعد البيانات، يقوى هذا الرابط أو يضعف بحسب مدى التوافق في البنية وآليات العمل بين نموذج البحث في المحرك ونماذج البحث في قواعد البيانات، فإذا ما كان التوافق في أعلى درجاته كان التفاعل بين المحرك وبين قواعد بيانات الويب في أفضل حالاته، ومن ثم تبلغ نسبة الدقة في النتائج المسترجعة أفضل مقاييسها، والعكس من ذلك صحيح.

١٩. تعكس نماذج البحث داخل محركات بحث العينة واقع تنظيم المعلومات داخل المحركات، إلى جانب طبيعة تنظيم المعلومات داخل مواقع الويب غير المرئية التي تخضع للبحث بواسطة محركات بحث العينة؛ فعندما يحتوي نموذج البحث المتعدد على خيارات البحث بالحقول الببليوجرافية، فإن ذلك يعني أن محرك البحث هذا يتعامل مع قواعد بيانات ذات بنية الجداول والتسجيلات، إلى جانب احتوائه على تسجيلات مكونة من حقول ببليوجرافية قابلة للبحث بواسطة المستفيد. على الجانب الآخر فإن عدم وجود خيارات البحث بالحقول الببليوجرافية يعني أن محرك البحث هذا يتعامل مع مواقع الويب التي لا تحوي تسجيلات ببليوجرافية أو تستخدم نظامًا مغايرًا لنظام الجداول والتسجيلات، وبذلك فهو يرتبط بالمواقع أكثر من ارتباطه بقواعد البيانات.

٢٠. توصلت الدراسة التجريبية إلى أن من بين محركات وأدلة بحث عينة الدراسة، كان أفضل محركات البحث هو incywincy، وأفضل أدلة بحث عينة الدراسة هو completplanet. وقد جاء ذلك من أن محرك البحث incywincy قد شمل مختلف الآليات التي يمكن أن يعمل بها محرك بحث الويب غير المرئية، على الرغم من جوانب قصور متعددة تم رصدها في أداء هذا المحرك. أما دليل البحث completplanet فقد خرج للويب داعمًا ومتوجهًا إلى التعامل مع المحتوى غير المرئي لقواعد بيانات الويب، ومن ثم شمل هذا

الدليل مجموعة كبيرة من آليات بناء أدلة الويب غير المرئية. وقد جاء دليلا قواعد البيانات findarticles ,magportal كأفضل أدوات بحث الدراسة تنظيمًا واسترجاعًا لمقالات ووحدات قواعد بيانات الدوريات الإلكترونية.

٢١. كشف غياب برنامج الزاحف في بعض محركات بحث عينة الدراسة عن حقيقة هذه المحركات في أنها لا تنتمي إلى محركات بحث الويب غير المرئية، وإنما يمكن تصنيفها على أنها أنواع أخرى من أدوات البحث ونظم الاسترجاع التي تحاول العمل على استرجاع المحتوى غير المرئي؛ لأن الأدوات التي لا تملك برنامج الزاحف لا تستطيع أن تجوب الويب باحثة عن نماذج وصفحات بحث قواعد البيانات أو مواقع الويب غير المرئية. وقد وصفت بعض أدوات بحث عينة الدراسة بوصف آخر غير أنها محرك بحث للويب غير المرئية ؛ حيث ذكرت أداة البحث infomine أنها مكتبة تخيلية virtual library للمصادر الرقمية على الويب، كما وُصفت أداتا البحث ,business sciencesearch عملهما بأنهما بوابة portal لمصادر المعلومات العلمية وإدارة الأعمال.

٢٢. إن برنامج الزاحف في محرك بحث الويب غير المرئية المقترح سيحمل عبء الوصول إلى نماذج البحث في واجهات مواقع وقواعد بيانات الويب غير المرئية، وذلك لتحليلها وتكشيفها، ومن ثم فإن اتجاهه في العمل سيكون على تتبع نماذج البحث وليس صفحات الويب ذات بنية html. فضلاً عن أن وصول برنامج الزاحف إلى نماذج بحث مواقع وقواعد بيانات الويب غير المرئية ليس هو الغاية في حد ذاته، وإنما تكمن الغاية في إمكانية تحليل وتكشيف هذه النماذج وتكشيفها؛ للتعرف إلى خصائص ومكوناتها وعناصر البحث التي تتركب منها. وهذا كله يهدف بدوره إلى فهم هذه النماذج من حيث آليات البحث بها وطبيعة عملها، حتى تتمكن محركات بحث الويب غير المرئية

مـن توجيـه استفسـارات البحـث إليهـا، واسـترجاع مجموعـات قواعـد البيانـات بواسطتها.

٢٣. إن أي مـن أدلـة بحـث العينـة لم يتبـع نظامًـا منطقيًـا مـن أنظمـة التقسـيم أو التصنيف الموضوعي للمعرفة البشرية، كخطط التصنيف المرتبة منطقيًا أو قوائم رؤوس الموضوعات المرتبة هجائيًا. وكان الاعتماد السائد على تقسيم عـام خـالٍ مـن أي منهجيـة أو منطقيـة في تقسـيم المصطلحات وترتيبهـا داخـل قائمة التصفح، واختلفت أدلة بحث العينـة في مستويات بنـاء هـذه التقسيمات العامـة؛ فمنهـا مـا توقـف عنـد حـد التقسيم العـام والانتقـال مباشـرة منـه إلى المواقع التي تغطـي موضـوعات هـذا التقسيم، ومنهـا مـا جعـل التقسيم مـدخلاً أوليًا لتغيير نمط التصفح من التصفح بالمصطلحات إلى تصفح المقالات وأعداد الدوريات كمـا حـدث في دليـل findarticles الـذي عمـل علـى تقسيم الـدوريات الرقمية داخله إلى أقسام موضوعية عامة، ثم تم ترتيب عناوين الدوريات تحت كل قسم ترتيبًا هجائيًا ثم رتبت أعداد كل دورية ترتيبًا زمنيًا بحسب تواريخ نشرها. ومن ثم يمكن القول إن التقسيم المتبع في دليل findarticles هو أفضل أنظمة الترتيب العامة المتبعة داخل أدلة عينة الدراسة.

٢٤. اعتمدت أدلة بحث عينة الدراسة على اتباع مستوى التكشيف العام لقواعد البيانات، وتمثل ذلك في أن أغلب النتائج المسترجعة لهذه الأدلة كان متمثلاً في استرجاع مواقع أو قواعد الويب غير المرئية ذاتها، دون القدرة على استرجاع محتوياتها الداخلية. ويدل ذلك على أن عمل أغلب أدلة البحث وقف عند حد الحصر والترتيب دون الوصول إلى حد التكشيف والتحليل للمحتوى غير المرئية لقواعد ومواقع الويب غير المرئية.

٢٥. جـاء اسـتخدام آليـات البحـث أكثـر في قواعـد البيانـات النـصية دون قواعـد بيانات الوسائط المتعددة، ويرجع ذلك إلى أن القواعد النصية تقترب أكثر من

المجالات العلمية المتخصصة، خلاف ما تكون عليه قواعد الوسائط المتعددة من عمومية المجال الموضوعي. أيضًا فإن القواعد غير المتخصصة موضوعيًا سواء أكانت نصية أم وسائط متعددة غالبًا ما تلجأ إلى الاعتماد على الشكل البسيط من آليات البحث بالاعتماد على توفير فراغ لكتابة كلمات البحث فقط، أما تلك التي تقترب من المجالات العلمية المتخصصة فإنها تتوقع أن يكون المستفيد على دراية بآليات الربط الرياضي بين الكلمات مثل آليات الربط البوليني.

٢٦. إن الويب، منذ أن تخطت حدود الملفات النصية إلى الأنواع الأخرى من الملفات، جعلت التسجيلات النهائية داخل قواعد بيانات الويب تنقسم إلى قسمين:

الأول : مجموعة الحقول البيليوجرافية المكونة لشكل التسجيلة النهائية داخل قاعدة البيانات وهي حقول بيانات الوصف.

الثاني : الملف الذي يحوي المعلومات النهائية، ويختلف في شكله من قاعدة إلى أخرى بحسب طبيعة المعلومات التي يحملها فهناك ملفات النصوص doc وملفات pdf التي تحوي البيانات النصية، أيضًا هناك ملفات wmv, png, mp3... وغيرها من الأنواع التي تحوي المعلومات المصورة والمسموعة.

٢٧. لقد أفرزت البيئة الرقمية الممثلة في الويب نوعًا أو أسلوبًا جديدًا من أساليب البحث عن المعلومات الرقمية، تمثل هذا الأسلوب فيما يسمى بأسلوب الاسترجاع المبني على المحتوى Content Based Retrieval CBR. وهو الأسلوب التقني المقابل لأسلوب البحث والاسترجاع المبني على الكلمات أو النصوص Text Based Retrieval TBR. وتكمن آلية عمل أسلوب البحث المبني على المحتوى في استخدام عناصر البنية texture والشكل shape واللون color لاسترجاع ملفات المعلومات المصورة، سواء تلك الثابتة أم المتحركة.

٢٨. مكنت التقنيات الرقمية قواعد بيانات الويب من القدرة على العمل بنمط تفاعلي ومتغير ؛ حيث يمكن لقواعد بيانات الويب عرض محتوياتها الداخلية الآن في واجهة واحدة وإن اختلفت المعلومات المعروضة في طبيعتها ، كما يمكن لقاعدة بيانات الويب صياغة المحتوى الداخلي وتغييره بشكل ديناميكي يمكن المستفيد من تجميع معلومات مختلفة والاطلاع على موضوعات متعددة في آن واحد. ولقد شكل المحتوى الديناميكي أو التفاعلي لمواقع الويب عامة حجر عثرة أمام محركات بحث الويب التي تجد نفسها أمام محتوى سريع التغيير، لا يمكن نسخه أو التحكم به ، وهذا العامل هو أهم عوامل ظهور الويب غير المرئية. وقد كشفت الدراسة عن ثمة رابط بين طبيعة قواعد البيانات التجارية أو الوسائط المتعددة التي تسعى لجذب المستفيد ومحاولة التميز بطرق العرض والإبهار أمام المستفيد ، خاصة إن كانت المعلومات المعروضة يطغى عليها التنوع في جوانبها الموضوعية أو أشكالها المصورة والنصية. وقد تمثل ذلك في قواعد بيانات الصوت rhapsody, allmusic ، وقواعد بيانات الأفلام movietome, world of movies ، إلى جانب قاعدة البيانات التجارية amazon.

٢٩. كشفت الدراسة عن أن مشكلة استرجاع محتوى الويب غير المرئية لا تكمن فقط في المشكلات التقنية أو الفنية التي تواجه محركات وأدلة بحث الويب غير المرئية ، وإنما تكمن في كثير من القضايا الجانبية المؤثرة والمتحكمة بدورها في بناء وتصميم المحتوى غير المرئي ؛ وهذه القضايا مثل القضايا التجارية الخاصة بملكية نشر أو بيع مواد المعلومات، أيضًا القضايا الأخلاقية المتعلقة بالتعامل مع محتوى يشمل الوسائط المتعددة وموضوعات تتفق مع طبيعة الثقافة العربية الإسلامية أو لا تتفق، بالإضافة إلى القضايا الأمنية التي تتعلق بحماية أمن وخصوصية المعلومات التي قد لا تقتصر على شركات عامة ، وإنما تمتد إلى القطاعات الحكومية العسكرية.

٣٠. قدمت الدراسة في فصلها السادس النتيجة النهائية الممثلة في مواصفات محرك بحث مقترح (STORM) يدعم بحث واسترجاع محتوى قواعد بيانات الويب غير المرئية. وقد أطلق الباحث على هذا النموذج المقترح مسمى STORM. ويرجع الباحث هذه التسمية المحددة إلى مجموعة من الأسباب هي: يعكس اسم العاصفة الحالة القوية التي يكون عليها محرك بحث الويب غير المرئية من جمع وضم محتوى ومجموعات مختلف قواعد بيانات الويب التي سيقابلها، أيضًا فإن الفعل "عصف" يدل على قوة محرك البحث المقترح في تحطيم كل الحدود وكسرها للدخول إلى محتوى وملفات قواعد بيانات الويب التي لم يكن لبرامج الزاحف الوصول إليها من قبل، هذا بالإضافة إلى أن العاصفة تحمل دائمًا في قلبها كل ما تجده من أشياء دون التركيز على شيء محدد، وهو ما سيقوم به محرك STORM المقترح في الدخول والتعامل مع ملفات قواعد البيانات النصية وغير النصية، كالصور والصوت والفيديو. وقد ضم هذا الفصل مختلف جوانب وعناصر بناء محرك البحث المقترح من خلال الحديث أنماط معالجة وتنظيم قواعد البيانات إلى جانب أنماط استرجاعها والتعامل مع ملفاتها، ثم رسم الصورة المقترحة لما سيكون عليه برنامج الزاحف في التعامل مع قواعد بيانات الويب غير المرئية.

ثانيًا: التوصيات:

على ضوء النتائج السابقة التي توصلت إليها الدراسة يقترح الباحث بعض التوصيات التي يوجزها فيما يلي، ومن خلال تقسيمها إلى الفئات الثلاث التالية:

(أ) توصيات خاصة بشركات تصميم أدوات بحث الويب:

١. ينبغي الاهتمام بتصميم برنامج الزاحف والتركيز على شروط التعامل مع قواعد البيانات داخل ملف عمل برنامج الزاحف robot.txt. حيث ستمثل البنية التقنية لبرنامج الزاحف الداعم الأول في تحقيق التوافق بين محركات بحث الويب غير المرئية وبين مجموعات قواعد بيانات الويب.

٢. ينبغي التغاضي عن تقسيم أدوات بحث واسترجاع الويب غير المرئية إلى محركات وأدلة، بل الاتجاه نحو بناء بوابات الويب غير المرئية، بما لها من قدرات عملية وبحثية تجمع بين آليات عمل كل من محركات وأدلة بحث الويب.

٣. ضرورة الوعي التام بالمتطلبات الفنية التقنية لبناء محركات بحث الويب غير المرئية، من حيث التركيز على فكرة التحليل والتكشيف لنماذج بحث واجهات مواقع وصفحات الويب، إلى جانب التعامل مع هذه النماذج طرحًا للاستفسارات وتلقيًا للنتائج المسترجعة.

(ب) توصيات خاصة بالباحثين والمتخصصين في دراسة الويب ونظم المعلومات:

١. ينبغي الاتجاه نحو دراسة الويب غير المرئية من كافة تفاصيلها وذلك لبناء الأساس التنظيري للويب غير المرئية. فضلاً عن بناء المجال والإطار الموضوعي الخاص بتناول التقسيمات والموضوعات الفرعية لظاهرة الويب غير المرئية.

٢. ينبغي على باحثي وطلاب الماجستير والدكتوراة في مجال المكتبات وتكنولوجيا المعلومات الالتفات إلى مجال المعلومات الرقمية وظواهرها المستحدثة التي تشكل مشكلات بحثية في تنظيم واسترجاع المعلومات الرقمية. وهي بذلك تمثل العمود الفقري لعلم المعلومات في شقيه مصادر المعلومات الرقمية، ونظم المعلومات على الويب.

٣. ينبغي على أقسام ومدارس المكتبات مسايرة الأنماط المستحدثة من ظواهر علم المعلومات عامة، والاهتمام بموضوع الويب غير المرئية، خاصة كمقرر يتم تدريسه بين مقررات علم المعلومات، لأن هذه الموضوعات أضحت تمثل واقعًا يفرض نفسه على مجال تنظيم واسترجاع المعلومات الرقمية.

(ج) توصيات خاصة بالهيئات والمؤسسات الفنية القائمة على دعم الويب وتشغيله:

١. ينبغي على هذه الهيئات التنويه وطرح الظواهر البحثية الجديدة التي تفرزها الويب في طريق تطورها وتحديثها، أمام الباحثين للعمل على استنباط واستنتاج المقترحات الملائمة لهذه الظواهر. وهو ما ينطبق على واقع الويب غير المرئية.

٢. ينبغي على هذه الهيئات السعي في فتح نوافذ للتفاعل مع الباحثين المختلفين على مستوى العالم، حتى يمكن مناقشة ودعم المقترحات الجديدة التي تخص تطوير أداء الويب وتحديثه. وهو ما كان سيفتح نافذة التفعيل والتطبيق أمام دراسة مثل دراستنا الحالية.

٣. ضرورة مساهمة الهيئات والمؤسسات في عمليات تطوير مواقع الويب المرئية وغير المرئية، وتسهيل إتاحة خدماتها للمستفيدين.

وقد انبثقت عن الدراسة الحالية بضع مجالات تحتاج إلى دراسات منهجية هي:

١. **محركات بحث المكتبات الرقمية على الويب** : دراسة لاسترجاع مجموعات المكتبات الرقمية بواسطة محرك البحث

٢. **تنظيم واسترجاع مواقع المعلومات التفاعلية (مواقع الويب غير المرئية)** : دراسة تحليلية.

٣. **محركات بحث الملفات الرقمية غير المعيارية Non HTML files retrieval** .



الملاحق

الملحق رقم (١)

قائمة المصطلحات والمختصرات الأجنبية والمقابل العربي

المقابل	المصطلح
واجهات عرض الويب	Application programmer interfaces
التكشيف بالتعيين	Assigned indexing
البحث المترابط	Associative search
تحليل خصائص نموذج البحث	Attribute Analysis
الكشف الذاتي عن قواعد البيانات	Automated discovery of invisible databases
المنطق البولياني	Bolean logoic
النتائج المقسمة موضوعيًا	Categorized result
لغة التحويل العامة	Common Gateway Interface (CGI)
الاسترجاع المبني على المحتوى	Content Based Video Retrieval
الزحف	Crawling
التقسيم الموضوعي للبيانات	Data clustering
التواصل وقاعدة البيانات	Database connectivity
تطبيع قاعدة البيانات	Database Normalization
خادمات قواعد البيانات	Database servers
واجهات قواعد البيانات	Databases interface
محركات بحث قواعد البيانات	Databases search engines
نظم إدارة قواعد البيانات	DBMS
نظم إدارة قواعد بيانات سطح المكتب	Desktop Database Systems
نظم إدارة قواعد بيانات سطح المكتب	Desktop platforms

المصطلح	المقابل
Digitization	عملية الرقمنة أو التحول إلى الشكل الرقمي
Domain name	مجال موقع الويب
DSN	اسم مصدر البيانات
Dynamic documents	صفحات الويب الديناميكية
Dynamic URLs	روابط ديناميكية
Dynamic Web	الويب التفاعلية
Dynamic web information	محتوى الويب المتغير
Dynamic Web Resources	موارد الويب الديناميكية
Extended tag language	لغات الترميز الممتدة
Fast CGI or extended CGI	إصدارات لغة التحويل العامة
Field searching	بحث الملفات
Foam fields	حقول نموذج البحث
Form Analysis	تحليل نموذج البحث
Form Classifier	تصنيف نموذج البحث
Form parser	محلل النموذج
Form values	القيم المدخلة لنموذج البحث
FTP	بروتوكول نقل الملفات
Global access	الوصول العالمي
Header	الملف الفاتح لتسجيل الحقول
Host web page	صفحة الويب المضيفة
HTML	لغة ترميز النص الفائق
HTML form	نموذج بحث الويب

المصطلح	المقابل
HTTP	بروتوكول النص الفائق
Hypermedia	الوسائط الفائقة
Hypermedia databases	قواعد بيانات الوسائط المتعددة
ICMP	بروتوكول التحكم برسائل الإنترنت
IMAP	بروتوكول إتاحة رسائل الإنترنت
Indexer	برنامج المكشف
Information Extraction (IE)	اقتباس المعلومات
Initial field set	القيم الذاتية لحقول البحث
inverted index	الكشاف المقلوب
Invisible web search engines	محركات بحث الويب غير المرئية
Invisible web	الويب غير المرئية
Invisible web search portal	بوابة بحث قواعد بيانات الويب
IP	بروتوكول الإنترنت
Interaction	التفاعلية
Java applets	معاملات لغة التحويل
Link Classifier	تصنيف الروابط الفائقة
Links	الروابط بين الصفحات
MARC	الفهرسة المقروءة آليًا
MARC Tags	حقول معيار الفهرسة الآلية (مارك)
Mediator	برنامج الوسيط لإدارة محرك قواعد البيانات
Memory sharing	المشاركة في الذاكرة
Meta Search Engines	محركات البحث المتعددة

المقابل	المصطلح
محارف وحقول لغات الترميز	Meta tag
البيانات عن البيانات أو البيانات الخلفية	Metadata
التحويل	Migration
عملية التصغير والتخزين في البيئة الرقمية	Minimization
نظم إدارة قواعد بيانات الوسائط	MMDBMS
نماذج لقواعد بيانات الخادم	MySQL, DB2, Paradox
نظم إدارة قواعد البيانات الشبكية	Network DBMS
العقد في النص الفائق	Nodes
نظم البيانات العلائقية متعددة الوسائط	Object/relational database management systems
مركز المكتبة المحسب على الخط المباشر	OCLC
قاموس مصطلحات المكتبات والمعلومات على الخط المباشر	ODLIS
أنظمة إدارة سطح المكتب	Office systems
لغة الويب الوجودية	Ontology Web Language
مصادر المعلومات المفتوحة	Open source
نظم التشغيل	Operating systems
رأس ملف صفحة الويب	page head
معاملات التغيير لروابط الويب	parameters
لغة كتابة الواجهات العامة CGI	perl
لغات تحويل الويب	PHP or perl programming
معيار التحقيق	precision
مبدأ السؤال المسبق واللاحق	pre-query and post query approach

المصطلح	المقابل
Real time search	البحث المتعدد في وقت واحد
Recall	معيار الاستدعاء
Reformulator	برنامج البحث في النتائج
Registration	الارتباط أو الاشتراك
Relational Database Management System	نظم إدارة قواعد البيانات العلائقية
Relevance	الصلة والارتباط بالموضوع
Required fields	قيم حقول البحث الإلزامية
Resolution	كثافة المواد المصورة
Response Analysis	تحليل النتائج المسترجعة
Response set	مجموعات النتائج المسترجعة
Result merger	مدمج النتائج
Robot.txt	ملف عمل برنامج الزاحف
Script	النصوص المكودة في الويب
Scripting language	لغات التحويل
Search box	فراغ البحث
Search engines	محركات البحث
Search labels	عناوين حقول البحث
Server Database Systems	نظم قواعد الخادمات
Site map	خريطة موقع الويب
Sites cookies	بطاقات تعريف المواقع
SMTP	بروتوكول نقل البريد الإلكتروني
Software licences	اتفاقيات تراخيص البرامج

المصطلح	المقابل
Spidering	عمل برنامج العنكبوت
spreadsheet	ورقة البيانات
Static content	المحتوى الثابت
Stop List	قوائم التوقف في التكشيف الإلكتروني
Structured Query Language (SQL)	لغة الاستفسارات المهيكلة
Subschemas	نماذج تخطيط فرعية
Subscription fee	الاشتراك بالمقابل المادي
Surface web	الويب السطحية " المرئية"
TBDR	استرجاع المحتوى المبني على النصوص
Thesaurus	المكانز
Three-tier Architecture	نظام الربط والبحث الثلاثي
Thumbnails	اللقطات
Two-tier Architecture	نظام الربط والبحث الثنائي
Unicode	التمثيل الموحد
Variable fields	الحقول المتغيرة في المحتوى
Web browser	متصفح الويب
Web directory	دليل الويب
Web navigation	الإبحار عبر الويب
Web servers	خادمات الويب
Web portal	بوابات الويب
Z39.50	معيار استرجاع المعلومات

الملحق رقم (٢)

أشكال توضيحية لمحركات بحث الويب غير المرئية(❖)

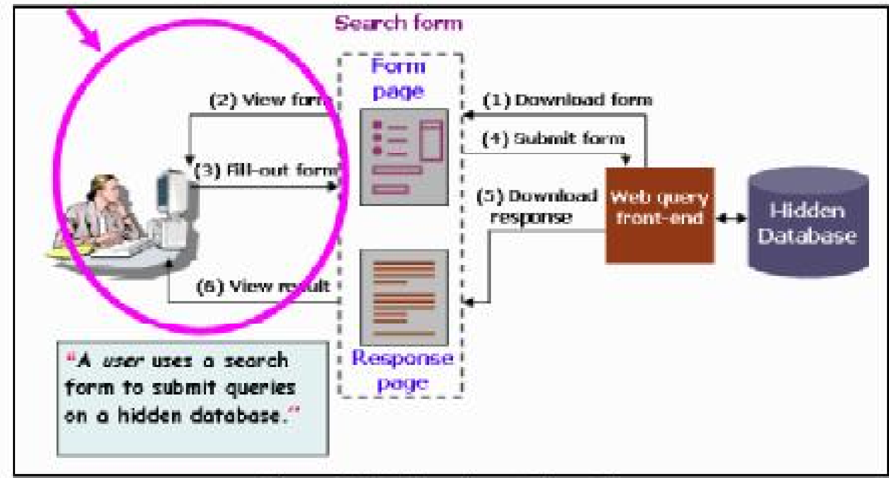


Figure 2 (a): User form interaction

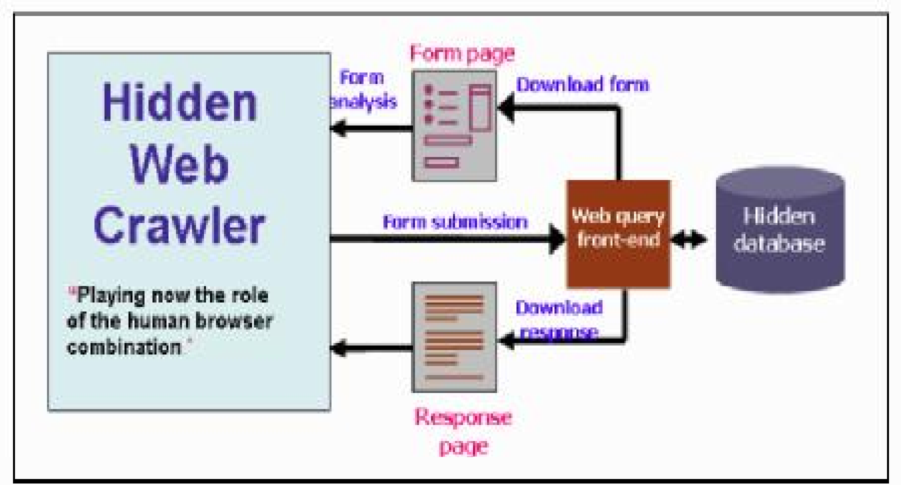


Figure 2 (b): Crawler form interaction

(❖) اقتبس الباحث هذه الاشكال اعتمادًا على أهم المصادر التي تناولت الويب غير المرئية، وهو :
El-Desouky I., Ali, Ali, Hesham A. and El-Ghamrawy, Sally M. A new framework for domain-specific hidden web crawling based on data extraction technieques, Mansoura University, Egypt, 2007, cited at 12/4/2009, cited at http://www.icict.gov.eg/ICICT-2006/Papers/Design%20for%20Information%20Retrieval%20Systems%20o

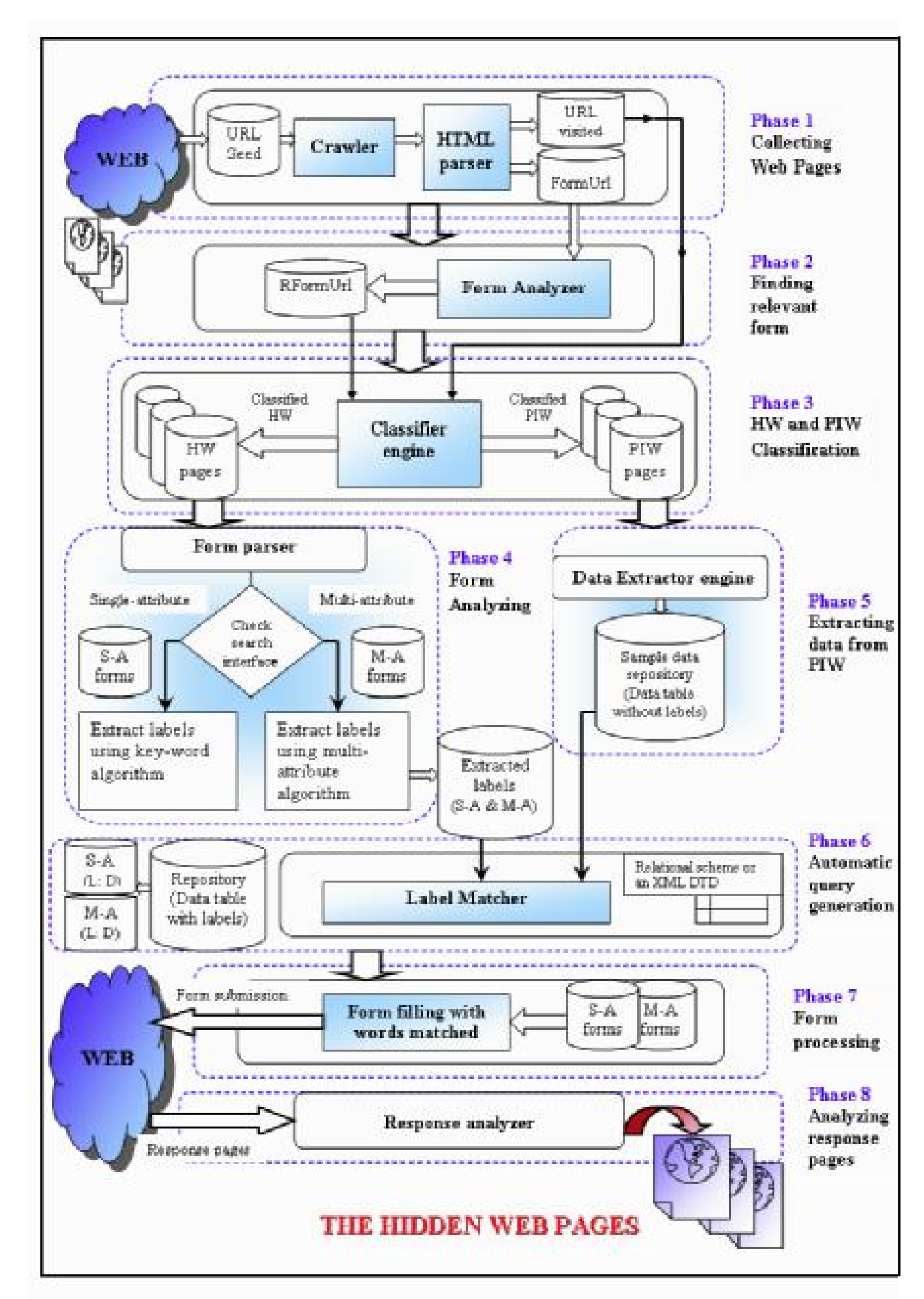


Figure 3: The proposed Hidden web crawler Framework

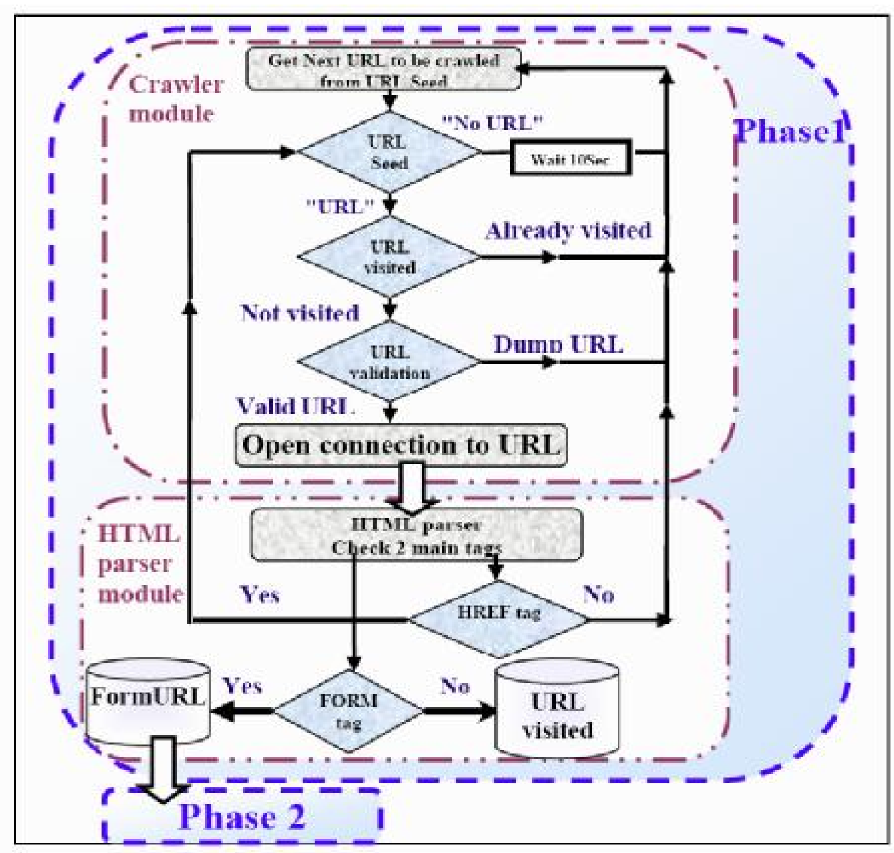


Figure 4: The proposed algorithm for phase1 (collecting web pages)

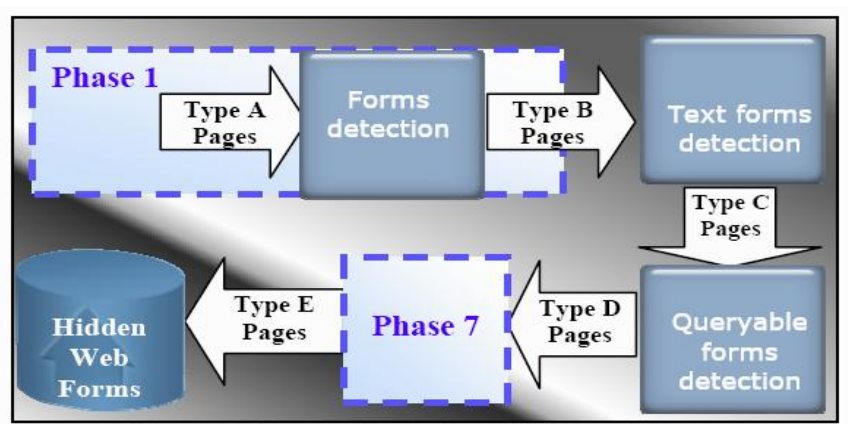


Figure 5: Stages of detecting hidden web forms

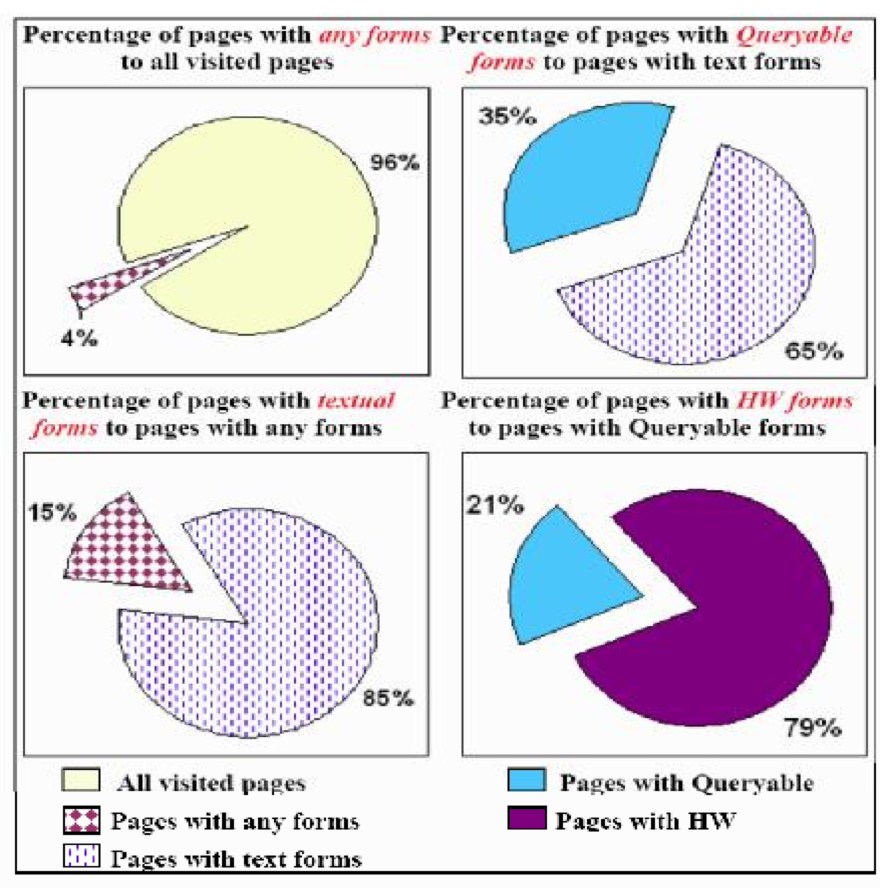


Figure 13: Analysis of web pages with forms

Author:

Title:

ISBN:

Search now

Figure 7(b): A multi-attribute search interface



Figure 10: Start window of the crawler (testing phase1)

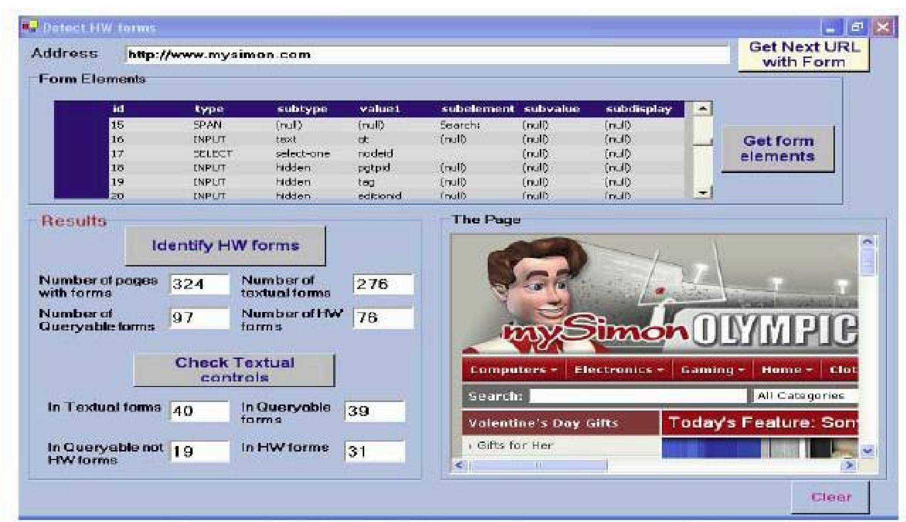


Figure 12: Check hidden web resources window

المصادر والمراجع

أولاً: المصادر العربية:

١. أحمد فرج أحمد. معيار Z39.50 : أداة أساسية لاسترجاع المعلومات عبر بوابات المكتبات .- cybrarians journal، ع ١١ (ديسمبر ٢٠٠٦م) ، تاريخ الاستشهاد ٢٠/١٠/٢٠٠٧م .- مسار :

http://www.cybrarians.info/journal/no11/z39.50.htm

٢. الإنترنت، مجلة المعلوماتية، ٢٠٠٠م، تاريخ الاستشهاد ٢٥/٨/٢٠٠٧م، مسار:

http://informatics.gov.sa/modules.php?name=Sections&op=viewarticle&artid=47

٣. البوابات(Portals) . مشروع سمو الشيخ محمد بن راشد آل مكتوم لتعليم تكنولوجيا المعلومات، ٢٠٠٥م، تاريخ الاستشهاد ٢٠/٧/٢٠٠٧م، مسار http://www.itep.ae/arabic/EducationalCenter/Articles/portal_01.asp

٤. سيد ربيع سيد. محركات بحث الصور الثابتة : دراسة تحليلية .- ط١.- الرياض : مكتبة الملك فهد الوطنية، ٢٠٠٧م .-ص ١٢٠.

٥. سيد ربيع سيد. محركات بحث الصور على الإنترنت : دراسة تحليلية لوضع مواصفات محرك بحث صور... ، إشراف هانئ محيي الدين عطية ، جامعة بني سويف، ٢٠٠٥م، أطروحة ماجستير، ص ٢٠.

ثانيًا: المصادر الأجنبية:

6. ACM Digital Library. New York (NJ) : Association for Computing Machinery. 2008, Cited at at: http://www.acm.org/dl/

7. ARMS, W. et al. An architecture for Information in Digital Libraries. D-Lib Magazine . February 1997. Cited at at: http://www.dlib.org/dlib/february97/cnri/02arms1.html .

8. Arroyo, Natalia. What is the invisible web? invisible web? A crawler perspective, Internet Lab, cited at http://cybermetrics.wlv.ac.uk/AoIRASIST/Arroyo.Invisible_web_crawlers.pdf

9. arXiv.org e-Print archive . Ithaca (NY) : Cornell University, 1991. Cited at at: http://arxiv.org/ .

10. Barbosa, Luciano. Combining Classifiers to Identify Online Databases, University of Utah, 2005, cited at 20/5/2009, cited at http://www2007.org/htmlpapers/paper603/

11. BARKER, Joe. Invisible Web: What it is, Why it exists, How to find it and Its inherent ambiguity. Finding Information on the Internet: A Tutorial, 2001 cited at 15.10.2005. Cited at at: http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html.

12. Beck, Susan E. Searching Web-Based Databases. Information Literacy, 2007,cited 30/6/2007cited at http://lib.nmsu.edu/resources/dbpqd.html

13. Berkhin, Pavel. Survey of Clustering Data Mining Techniques. Accrue Software, Inc., 2003, cited 20/9/2007, cited at http://www.ee.ucr.edu/~barth/EE242/clustering_survey.pdf

14. CARPENTER, Leona et al. OAI for Beginners - The Open Archives Forum online tutorial . S.l : Open Archives Forum, c2003, Cited at 2005-10-03. at: http://www.oaforum.org/tutorial/ .

15. CHANG, K. C.-C.; HE, B.; LI, C.; PATEL, M.; ZHABG, Z. Structured databases on the web: observations and implications. In ACM SIGMOD Record, Volume 33 , Issue 3 (September 2004). New York, NY, USA : ACM Press, 2004 cited at 13.06.2005.Cited at at: <http://doi.acm.org/10.1145/1031570.1031584>

16. Chapple , Mike. What is a Web Database. About, Inc, 2007, citd 15/6/2007, cited at http://databases.about.com/od/specificproducts/

17. Chapple , Mike.Choosing a Database for Your Organization.2005, cited at 10/10/2008, cited at http://databases.about.com/od/administration/a/choosing_a_db.htm

18. Chapple, Mike. Database Normalization basics, about.inc, 2006, cited at 7/7/2007, cited at http://databases.about.com/mbiopage.htm

19. Cohen , Laura. Understanding the World Wide Web. Internet tutorials, 2007, cited 20/6/2007, cited at http://www.internettutorials.net/

20. Cohen, Laura. The Deep. internet tutorials, 2006, cited 20/6/2007, cita at http://www.internettutorials.net/deepweb.html

21. Database Administrator Jobs and information. UniXL , 2004, cited 2/6/2007, cited at http://www.unixl.com/dir/information_technology/databases/

22. Databases. Find my hosting, 2002, cited 10/7/2007, cited at http://www.findmyhosting.com/index.asp

23. Dave Hillman. Relational Databases , the author, 2005,cited 17/2007, cite at http://dhillman.com/theplace/databoot/databases.htm

24. Doe, Charles G. Web-Based Databases and Search Tools, Information Today, Inc, 2007, 30/6/2007, cited at http://www.infotoday.com/MMSchools/sep04/doe.shtml

25. El-Desouky I., Ali, Ali, Hesham A. and El-Ghamrawy, Sally M. A NEW FRAMEWORK FOR DOMAIN-SPECIFIC HIDDEN WEB CRAWLING BASED ON DATA EXTRACTION TECHNIQUES, Mansoura University, Egypt, 2007, cited at 12/4/2009, cited at

http://etd.gsu.edu/theses/available/etd-07272007-215621/unrestricted/desai_lovekeshkumar_200707_ms.pdf

26. E-PRINT Network : Research Communication for Scientists and Engineers. Oak Ridge (TN, USA) Dept of Energy, Office of Scientific and technical Information.2008?, Cited at at: http://www.osti.gov/eprints/ .

27. ETRDL : ERCIM Technical Reference Digital Library . Brno (CZ) : Fakulta informatiky Masarykovy univerzity.2008?, Cited at at: http://dienst.muni.cz/ .

28. ExLibris. MetaLib Overview . cited at 2005-10-03. Avaliable at: http://www.exlibrisgroup.com/metalib.htm .

29. Google and the deep web. Blogger, 2007, cited 1/7/200٨, cited at http//glinden.blogspot.com/2007/03/google-and-deep-web.html

30. GRAVANO, LUIS and IPEIROTIS, PANAGIOTIS G. QProber: A System for Automatic Classification of Hidden-Web Databases, Columbia University, 2003, cited 1/8/2007, cited at htt//qprober.cs.columbia.edu/publications/tois2003.pdf

31. Hai He and Meng, Weiyi. Constructing Interface Schemas for Search Interfaces of Web Databases, 2005, cited 208/2007, cited at www.cs.binghamton.edu/~meng/pub.d/He_p136.pdf

32. Hamilton, Nigel. The Mechanics of a Deep Net Metasearch Engine, Turbo10 Metasearch Engine, 2004, cited 15/10/2007, cited at http://www2003.org/cdrom/papers/poster/p170/poster/poster.html

33. HE, B.; CHANG, K.C.-C. Making Holistic Schema Matching Robust: A Ensemble Approach. In Proceedings of the 2005 ACM SIGKDD Conference (KDD 2005) (Full Paper), Chicago, Illinois, August 2005.

Cited at at: http://eagle.cs.uiuc.edu/pubs/2005/robustmatching-sigkdd05-hc-jun05.pdf

34. HE, B.; PATEL; M.; ZHANG, Z.; CHANG, K. C.-C. Accessing the Deep Web: A Survey. Technical Report, Department of Computer Science, UIUC, July 2004. cited at 17.10.2005] Cited at at: http://eagle.cs.uiuc.edu/tr/dwsurwey-tr-hpzc-jul04.pdf

35. HE, B.; ZHANG, Z.; CHANG, K. C.-C.; MetaQuerier: Querying Structured Web Sources On-the-fly. In Proceedings of the 2005 ACM SIGMOD Conference (SIGMOD 2005), System Demonstration, Baltimore, Maryland, June 2005 cited at 20.10.2005]. Cited at at: http://eagle.cs.uiuc.edu/pubs/2005/querytranslationdemo-sigmod05-hzc-mar05.pdf .

36. He, Hai and Yu, Clement. Automatic Extraction of Web Search Interfaces for Interface Schema Integration, 2002, cited 11/11/2007, cited at www.cs.binghamton.edu/~meng/pub.d/WWWposterhe.pdf

37. Hedley,Yih-Ling and Younas, Muhammad. Sampling, Information Extraction and Summarisation of Hidden Web Databases, School of Mathematical and Information Sciences, Coventry University, 2006, cited 15/9/2007, cited at http://portal.acm.org/citation.cfm?id=1225816.1225818

38. Hillman, Dave. Relational Databases , the author, 2005, http://dhillman.com/theplace/databoot/reldbs.htm

39. Hobbins, Katie. Surfing the Invisible Web, 2001, cited at 10/10/2008, cited at www.cs.binghamton.edu/~meng/pub.d/WWWposterhe.pdf

40. Information Retrieval Models. (SCILS) at Rutgers University, 2002, cited at 1/8/2007, cited at Cited at http://www.scils.rutgers.edu/~aspoerri/InfoCrystal/Ch_2.html

41. Introduction to Databases for Web Developers : What is a Database. Extropia, 2007, cited 25/5/2007, cited at http://www.extropia.com/tutorials/sql/what_is_database.html

42. Introduction to Web Databases. 2002 by Scott Nicholson. http://www.askscott.com/sec2.html

43. Invented by Alon Y. Halevy, Jayant Madhavan, and David H. Ko
US Patent Application 20060230033
Published October 12, 2006
Filed on April 5, 2006 http://appft1.uspto.gov/netacgi/nph-Parser?

44. Invisible web. CloserLook Search Services Inc. 2002, cited 5/8/2007, cited at http://www.searchengineguide.com/wi/2002/0327_wi2.html

45. invisible web. Closerlook, 2006, cited at 25/7/2007,cited at http://www.closerlooksearch.com/invisibleweb.aspx

46. Ipeirotis, Panagiotis G. and Gravano, Luis. Distributed Search over the Hidden Web: Hierarchical Database Sampling and Selection, Columbia University, 2002, cited 10/8/2007, cited at http://qprober.cs.columbia.edu/publications/vldb2002.pdf

47. Ke-Thia Yao, Robert Neches. Dynamic Coordination of Information Management Services for Processing Dynamic Web Content, Information Sciences Institute, University of Southern California, 2005, cited 25/5/2007, cited at http://www2002.org/CDROM/refereed/613

48. KOEHLER, Wallace. Digital Libraries and World Wide Web Sites and Web Page Persistence. In Information Research 4.4, 1999 cited at 15.10.2005. Cited at at: http://informationr.net/ir/4-4/paper60.html .

49. Lauridsen, Helle. Scientist in the Electronic Library-is Deep Indexing of Use for Research?cites , cited at http://www.ffos.hr/lida/datoteke/LIDA2007-Lauridsen_monday.ppt

50. Lazar, Z. Peter and Holfelder, Peter. Web Database Connectivity with Scripting Languages O'Reilly & Associates, 1997,cited 25/6/2007, cited at http://www.oreilly.com/catalog/wj6/

51. LEWANDOWSKI, Dirk. Web Information Retrieval : Technologien zur Informationssuche im Internet. DGI: Frankfurt am Main, 2005. ISBN 3-925474-55-2. Cited at also at: http://www.durchdenken.de/lewandowski/web-ir/.

52. Lewandowski, Dirk. Web searching, search engines and Information Retrieval, heinrich-Heine-University, 2005, cited at 10/1/2009, cited at http://www.durchdenken.de/lewandowski/doc/isu_preprint.pdf

53. Li, Xian and Meng, Weiyi. EasyQuerier: A Keyword Based Interface forWeb Database Integration System, Renmin University of China, 2007, cited 10/9/2007, cited at http://www.springerlink.com/index/jx4p4572l707k541.pdf

54. Lu, Yiyao and Meng, Weiyi. Evaluation of Result Merging Strategies for Metasearch Engines, University of Illinois at Chicago, 2005, cited 20/10/2007, cited at http://www.cs.binghamton.edu/~meng/pub.d/Lu_p211.pdf

55. Luevane, Marsha and Chervenak, Joe. Web Searching 101: Search Tools, Strategies, and Tips, National Renewable Energy Laboratory, 2003, cited at 14/5/2009, cited at http://www-conf.slac.stanford.edu/interlab99/program/chervenak-luevane/searching101.ppt

56. Marc Grange.databases design, the author, 2006, cited 5/7/2007, cited at http://www.marc-grange.net/BDConception.htm

57. Meng, Weiyi and Yu, Clement. Building Efcient and Eective Metasearch Engines, State University of New York, , cited 28/8/2007, cited at http://portal.acm.org/citation.cfm?id=505284

58. metasearch engines. Clixmarketing, 2008, cited 10/10/2008, cited at www.clixmarketing.com/definitions.htm

59. metasearch.Blue Angel Technologies, Inc.2005, cited 25/9/2007, cited at http://www.BlueAngelTech.com

60. Multimedia DatabaseManagement Systems.ppt, 2008, Cited 30/6/2008, cited at http://www.chu.edu.tw/~leecc/962/B102%20and%20Databases.ppt

61. MYLLYMAKI, J. Effective Web Data Extraction with Standard XML Technologies. In Conference proceedings of the Tenth International World Wide Web Conference May 1-5, 2001 Hong Kong. 2001. 2001 . cited at07.10.2005] Cited at at: http://www10.org/cdrom/papers/pdf/p102.pdf .

62. NILSSON, Daniel ; NORIN, Henrik. Adaptive QoS Management in Dynamically Reconfigurable Real-Time Databases. NDLTD (2005-11-18) cited at20.11.2005]. Cited at ahttp://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-2800.

63. Nordic Metadata Project DC Metadata Viewer http://www.lub.lu.se/dc/nmd_viewer.pl .

64. Ntoulas, Alexandros. Crawling and Searching the HiddenWeb, supervising Gregory H. Leazer, UNIVERSITY OF CALIFORNIA, 2006, cited 208/2007, cited at http://www.dia.uniroma3.it/~vldbproc/017_129.pdf

65. Ntoulas, Alexandros. Crawling and Searching the HiddenWeb, University of California, 2006, cited at 20/5/2009, cited at http://portal.acm.org/citation.cfm?id=1061951.1061952

66. NTOULAS, Alexandros; ZERFOS, Petros; CHO, Junghoo. Downloading textual hidden web content through keyword queries. In International Conference on Digital Libraries Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, Denver, CO, USA New York : ACM Press, 2005, p. 100-109. Also cited at at: http://doi.acm.org/10.1145/1065385.1065407 Database content retrieval – searching in heterogeneous databases

67. ODLIS —Online Dictionary for Library and Information Science. Cited 10/5/2007, cited at http://lu.com/odlis/odlis_w.cfm

68. O'Neill, Edward T. Trends in the Evolution of the Public Web, D-Lib Magazine, April 2003, cited 20/6/2007, cited at http://www.dlib.org/dlib/april03/lavoie/04lavoie.html

69. online DB and searching.2004, cited 15/7/2007, cited at http://web.simmons.edu/~benoit/LIS530/602slides/17online.htm

70. Parts of a Database. Thinkquest, 2001, http://library.thinkquest.org/C0115420/Cyber-club%20800x600/DatabaseS/Parts%20of%20a%20Database.htm

71. Practical database design, IBM, 2003, http://www.ibm.com/developerworks/web/library/wa-dbdsgn1.html

72. Preservation Risk Management for Web Resources. D-Lib Magazine, 2002, cited 25/5/2007, cited at http://www.dlib.org/dlib/january02/kenney/01kenney.html

73. PRICE, Gary; SHERMAN, Chris. The Invisible Web : Uncovering Information Sources Search Engines Can't See. CyberAge Books, July 2001, p. 130-132. ISBN 0-910965-51-X. Wikipedia http://en.wikipedia.org/wiki/Dynamic_web_page

74. Raghavan, Sriram. Crawling the Hidden Web & Other Web-related Research, stanford university, 2004, cited 22/8/2007, cited at http://www.acm.org/fcrc/PlenaryTalks/Garcia-Molina.pdf

75. Reitz, Joan M. directory, ODLIS — Online Dictionary for Library and Information Science, Libraries Unlimited, 2006, cited 20/8/2007, cited at http://lu.com/odlis/odlis_d.cfm#directory

76. --------------------. metasearch, ODLIS — Online Dictionary for Library and Information Science, Libraries Unlimited, 2006, cited 20/9/2007, cited at http://lu.com/odlis/odlis_m.cfm# metasearch

77. ROCCO, Daniel; CAVERLEE, James; LIU, Ling; CRITCHLOW, Terence. Exploiting the deep web with DynaBot: matching, probing, and ranking. In International World Wide Web Conference, Special …, 2005. ACM Digital Library cited at 13.06.2005] Cited at at:http://doi.acm.org/10.1145/1062745.1062925 .

78. Ru, Yanbo and Horowitz, Ellis.Indexing the invisible web: a survey, University of Southern California, 2005, cited at 20/7/20007, citead at http://libraries.ou.edu/eresources/reserves/readings/7_31_2006_6_40_27_PM.pdf

79. Scammell, Lorna. Design your own database, the author, 2003, http://www.campus.ncl.ac.uk/databases/design/design.htm

80. Search Engine Indexing For Deep Web Pages. AccuraCast Limited, 2007, cited 1/8/2007, cited at http://www.accuracast.com/seo-weekly/invisible-web.php

81. SHERMAN, Chris. The Invisible Web. FreePint. 64. 2000 cited at 17.10.2005. Cited at at: http://www.freepint.com/issues/080600.htm#feature

82. Shestakov, Denis and Bhowmick, Sourav S. DEQUE: querying the deep web, Nanyang Technological University, Singapore, 2004, cited at 2/4/2009, cited at www.sciencedirect.com

83. Smith , Joan A., McCown, Frank and Nelson, Michael L. Observed Web Robot Behavior on Decaying Web Subsites, D-Lib Magazine, 2006, citd 1/9/2007, cited at http://www.dlib.org/dlib/february06/smith/02smith.html

84. -------------------. Testing the Surf: Criteria for Evaluating Internet Information Resources." The Public-Access Computer Systems Review 8, no. 3 (1997). Cited 20/5/2007, cited at http://info.lib.uh.edu/pacsrev.html.

85. Sol, Selena. Introduction to Databases for the Web, 1998, cited 10/6/2007, cited at http://www.databasejournal.com/sqletc/article.php/26861_1431601_1

86. Sriram Raghavan. Crawling the HiddenWeb, Stanford University,2001, cited 25/8/2007, cited at http://www.springerlink.com/index/jx4p4572l707k541.pdf

87. SUBJECT DIRECTORIES. Debbie Flanagan, 2004, cited 10/6/2007, cited at http://www.learnwebskills.com/search/subject.html

88. Sullivan , Danny . Search Engine Optimization & Marketing Glossary, Submitawebsite, Inc , 2004 , cited at 12/6/2008 , http://www.submitawebsite.com/seo_glossary.html

89. Summary of a New Client/Server/Database Architecture.2003, cited 25/6/2007, cited at http://www.semibase.com/database.html

90. Surface Web. Wikipedia : the free encyclopedia, 2007, cited 20/6/2007, cited at http://en.wikipedia.org/wiki/ Surface_ Web

91. System analysis and design ; chapter 7, 2008, http://aslnx1.uah.edu/~floydj/MIS497/chapter07.ppt

92. The Active Database By Cameron O'Rourke . oracle , 2003, http://www.oracle.com/technology/index.html

93. Timeless Principles of Design (Web Techniques, CMP Technology,2007,

http://www.ddj.com/dept/architect/184414356?pgno=1

94. Walny, Jagoda. SemaForm: Semantic Wrapper Generation for Querying Deep Web Data Sources, Supervisor: Dr. Denilson Barbosa

95. WARNICK, Walter L. et al. Searching the Deep Web: Directed Query Engine Applications at the Department of Energy. In D-Lib Magazine, January 2001, Vol. 7, Nr. 1. cited at 5.10.2005]. Cited at at: http://www.dlib.org/dlib/january01/warnick/01warnick.html .

96. Watson, James. Information Models for Web Structure, published by NIVA Inc. cited 20/5/2007, Cited at http://www.writersblock.ca/spring1998/feature.htm

97. Web Databases. MDA, 2002,cited 20/6/2007, cited at http://www.mda.org.uk/webdatabases.htm

98. web designs. Realife web designs, 2005, cited 15/5/2007, cited at http://www.realifewebdesigns.com/web-resources/web-tips.html .

99. Web Ignite corporation Indexing Deep Web Content, Visibility Factor, 2002, ited 1/8/2007, cited at http://www.visibilityfactor.com/articles/indexing_the_deep_web.html

100. Web protocols. Securepaynet, 2006, cited 10/5/2006, Cited at http://www.realifewebdesigns.com/web-resources/web-protocols.html.

101. Web-based Databases:Technical Implementation.ppt .cited 30/6/2007, cited at \ http://foodafrica.nri.org/enterprises/enterprisespapers/GhirmaMogesQuality.doc

102. Website databases and programming. DianeV. Web Design Studio, 2007,cited 20/6/2007, cited at http://dianev.com/internet-marketing.html

103. Website Indexing. Website-Designer, 2007, cited at 6/6/2007, cited at http://www.website-designer.net/main.php

104. What are the characteristics of a web-based database design? Blurtit.inc, 2007, cited 2/7/2007, cited at http://www.blurtit.com/ask.php

105. What is a Search Engine? Metamend.inc , 2004 , Visited at 12/6/2004 , http://www.metamend.com/search-engine-map.html

106. Woodcroft, Michelle and Fletcher, Janet. Establishing guidelines for the effective evaluation of web-based periodical bibliographic and full-text databases search interfaces, Southern Cross University, 2000, cited 1/11/2007, cited at www.vala.org.au/vala2000/2000pdf/Woo_Fle.PDF

107. Wu ,Wensheng WebIQ: Learning from the Web to Match Deep-Web Query Interfaces, University of Illinois,2006, cited at 1252009, cited at http://www.cs.binghamton.edu/~meng/pub.d/icde06-webiq.pdf

108. Z. Peter Lazar, Peter Holfelder. Web Database Connectivity with Scripting Languages O'Reilly & Associates, 1997, http://www.oreilly.com/catalog/wj6/

ثالثًا: مواقع الويب:

1. http://animalsearch.net/
2. http://business.exploritnow.com/
3. http://catalog.mpl.org.eg/
4. http://infomine.ucr.edu
5. http://medlineplus.gov/
6. http://movietome.com/
7. http://turbo10.com
8. http://www.academicinfo.net/
9. http://www.allmusic.com/
10. http://www.amazon.com/
11. http://www.beaucoup.com/
12. http://www.biography.com/
13. http://www.completeplanet.com/
14. http://www.egypt.gov.eg
15. http://www.eric.ed.gov/
16. http://www.genealogyspot.com/
17. http://www.incywincy.com/
18. http://www.nasa.gov/multimedia/imagegallery/index.html
19. http://www.rhapsody.com
20. http://www.scienceresearch.com/search/
21. http://www.scirus.com/srsapp/
22. http://www.world-of-movies.com/

23. http://wwwlib.umi.com/dissertations
24. www.corbis.com
25. www.findarticles.com
26. www.freefoto.com
27. www.gooogle.com
28. www.ifla.org
29. www.magportal.com